# 基于招聘数据集的求职指南

## 一、问题描述

本数据分析报告旨在利用包含大量招聘信息的数据集进行深入的数据分析和可视化展示，并提出相应的建议和策略以帮助求职者了解目前招聘市场的形式，以及帮助他们更好的找到适合自己的工作。

经过对数据集的观察，认为需要进行数据分析的主要内容包括：

**1. 市场概览**

这部分提供了对于当前招聘市场的整体情况的宏观视角。通过深入挖掘各种关键指标，求职者可以明确地了解各大城市和行业的招聘需求，以及公司的类型和规模。此外，通过分析学历要求，求职者可以更好地了解自己的竞争力和潜在机会。

——发布招聘信息最多的前六个城市饼状图

——发布招聘信息最多的前六个行业饼状图

——公司类型分布饼状图

——公司规模饼状图

——学历要求饼状图

——各个学历的招聘信息的数量

——各个学历的招聘人数之和

**2. 职位分析**

这部分可以:①帮助求职者了解特定职位的需求情况;②帮助求职者了解各种岗位对于学历的需求，以便求职者通过自己的学历确定想要从事的岗位，也便于求职者通过自己想要从事的岗位，了解其要求的学历;③帮助求职者了解各个城市对于岗位的需求，以帮助求职者通过自己所在的城市，判断有哪些岗位需求量大，也可以帮助求职者通过自己想从事的岗位，选择该岗位需求量大的城市

——岗位需求与学历关系分析 (以律师,软件工程师,教师,服务员为例)

——不同城市的岗位需求分析 (以律师,软件工程师,教师,服务员为例)

**3. 待遇与福利分析**

这部分专门针对公司提供的待遇和福利进行深入的分析。求职者可以根据这些数据，评估不同行业和公司提供的福利，并基于此做出决策。

——公司待遇分析

——不同行业五险一金的比例

——不同行业绩效奖金的比例

——不同行业专业培训的比例

——不同行业年终奖金的比例

——不同行业节日福利的比例

——不同行业带薪年假的比例

——高薪酬是否意味着更好的待遇？用五险一金,专业培训,带薪年假为例

**4. 经验与薪酬关系**

经验往往是决定薪酬的一个重要因素。这部分将帮助求职者了解，在相同的行业和职位下，经验如何影响他们的薪酬。

——经验对薪酬的影响

**5. 招聘信息完整度的选择**

求职者面对大量的招聘信息，有些信息是完整的，有些招聘信息并不完整。本部分根据每条数据的空值数量先定义了信息完整度的概念，通过分析信息完整度与招聘信息其它特点的关系，帮助求职者确定选择完整的招聘信息完善的工作还是不太完善的工作。

——提供招聘信息的完整度和质量方面的指标。

——招聘信息完整度分析

——招聘信息完整度与薪酬的关系

——招聘信息完整度与招聘人数的关系

**6. 行业与城市选择**

这部分旨在为求职者提供行业和城市的指南，以便他们根据自己的兴趣和经验选择合适的工作领域和地点。数据可以揭示哪些城市和行业为初入职场的求职者提供了最多的机会。

——哪些行业为在校生和应届生提供更多的机会

——哪些城市为在校生和应届生提供更多的机会

——总招聘人数前六的城市

——总招聘人数前六的城市的平均薪酬

——前六行业与在各城市的招聘总数的前六名

**7. 公司规模选择**

考虑公司的规模在求职过程中是十分重要的。这部分的数据分析会帮助求职者了解不同规模的公司在薪酬和待遇上的差异，从而做出更明智的决策。

——公司规模与薪酬的关系

——公司规模与待遇的关系

**8.公司类型选择**

不同类型的公司，例如国企还，民营公司，它们都有各自的工作文化和福利待遇。这部分的分析将帮助求职者更深入地了解这些差异，为他们的职业规划提供宝贵的建议。

——公司类型与薪酬的关系

——公司类型与待遇的关系

接下来，本文将使用R语言,借助Rstudio,围绕上面的问题，进行相应的分析，并得出相关的结论，以帮助求职者了解和寻找工作。

## 二、数据预处理

为了完成上面的数据分析，需要用到薪酬，地区，招聘人数，学历，经验等数据，但是在原始数据集中，存在着诸多问题，例如薪酬的数据格式不一，有的采用区间形式，有的采用一个值，单位也有元，千，万，薪酬发放频率也不一致，有年，月，天，时。因此，为了更好地进行后续的数据分析，需要先统一地进行一次数据预处理操作。

下面，将主要按照原数据集中字段的顺序展开序数

**1.由于数据量过于庞大，随机抽取原始数据集的1/100进行后续的分析**



**2.“薪酬”列**

2.1提取并计算中间值，将所有的区间都用中位数来代替。这样处理比较符合尝试，能够更好地反应薪酬水平。

2.2转换元，千，万等货币单位

2.3将所有的薪酬都转换为年薪

2.4为data新建一列，并将转化后的薪酬储存在其中，命名为薪酬\_processed,然后对比薪酬和薪酬\_processed,查看是否成功完成转换

2.5删除薪酬列,并将薪酬\_processed列命名为薪酬

**3.“地点及需求”列**

该列储存了大量有用的数据，但是都集中在这一列导致数据无法方便的使用，因此有必要将其中有用的数据进行分别提取

3.1提取地区信息

3.2提取经验信息

3.3提取在校生/应届生信息

3.4提取学历信息

3.5提取招聘人数

3.6提取发布时间

3.7删除地点及需求列

**4.“地点列”**

地点列现在储存的数据也并不规范，有的数据为城市，有的数据是城市-区域

4.1拆分地点列为城市和地区两列

处理后的数据集的部分列如下所示，其他未展示的列都是原来的样子：



**5.处理空值**

虽然数据中有很多缺失值,但是因为后续涉及到信息完整度这个概念,所以对空值不做处理

## 三、可视化展示与分析建议

1. 市场概览

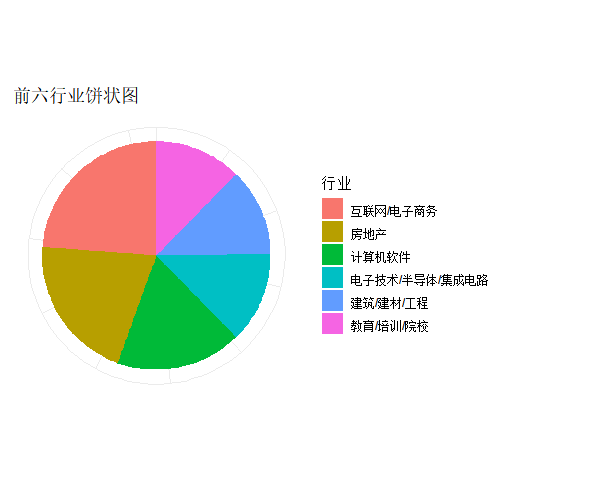
这部分旨在为求职者和应届生提供整体市场的趋势和情况。

1.1发布招聘信息最多的前六个城市饼状图

Rplot

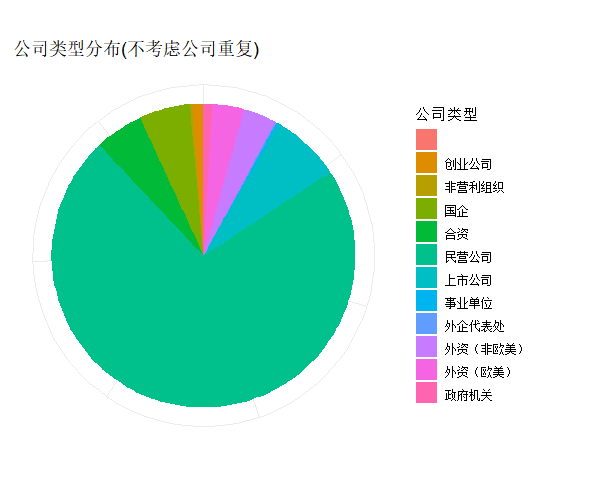
数据显示深圳,广州,北京,上海,杭州,武汉发布的信息数量较多,反映了他们有着相对较多的就业机会

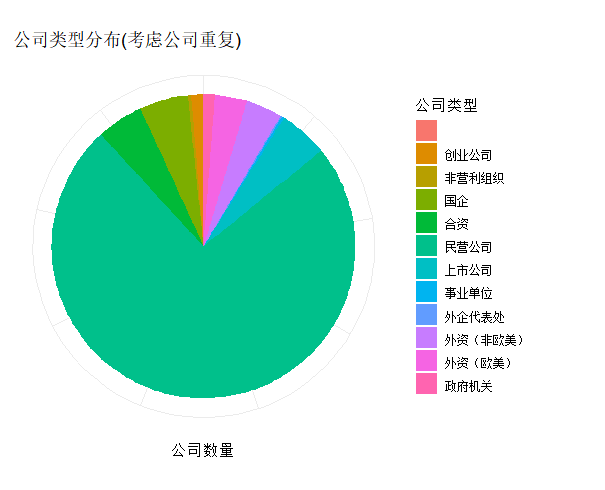
1.2发布招聘信息最多的前六个行业饼状图



数据发布招聘信息数量较多的六个行业,反映了这些行业有着相对较多的就业机会

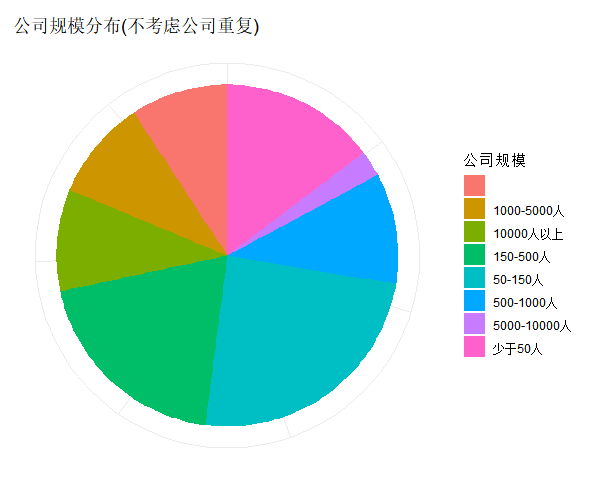
1.3公司类型分布饼状图

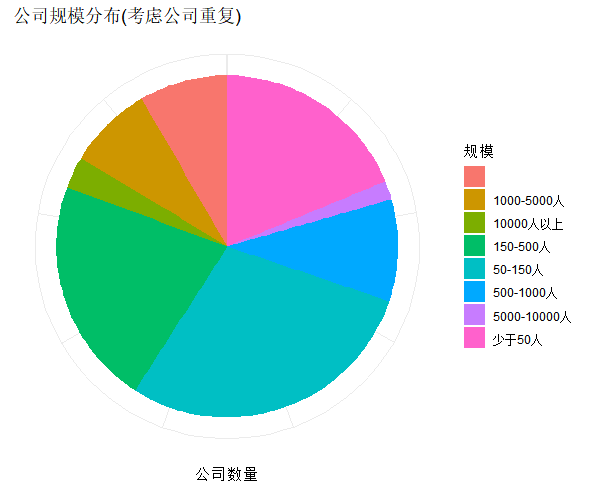




该图反映了发布招聘信息的公司的类型分布,显示了如今对劳动力需求的公司类型分布

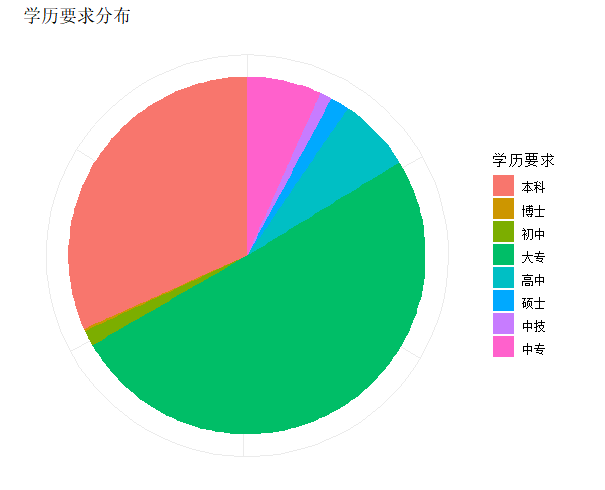
1.4公司规模饼状图





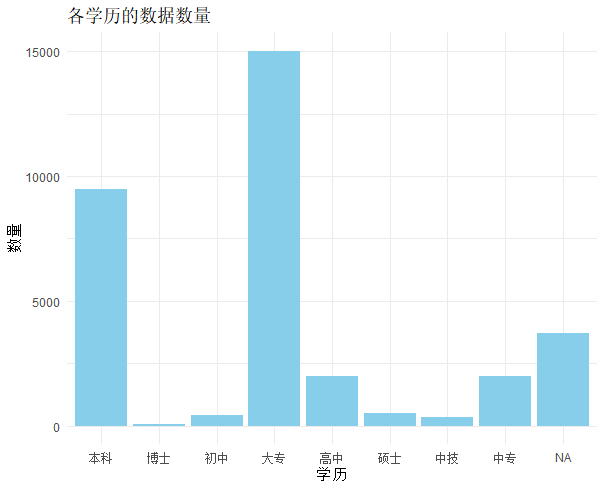
该图反映了发布招聘信息的公司的类型分布,显示了如今对劳动力需求的公司类型分布

1.5各个学历的招聘信息的数量饼状图

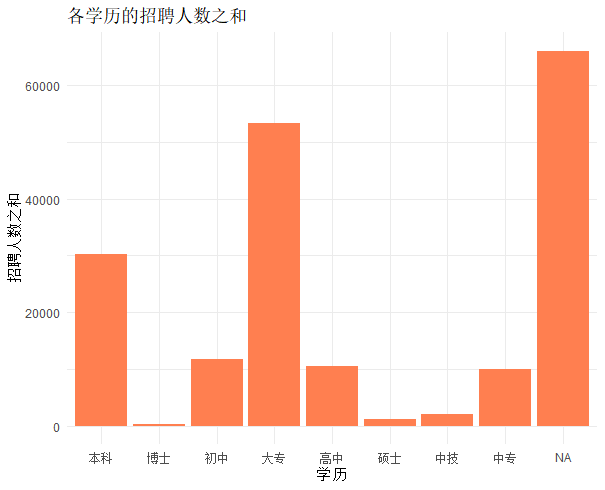


大专和本科是招聘信息中常常提到的最低学历要求,因此拿到本科和大专学历对于求职十分有利

1.6各个学历的招聘信息数量柱状图



1.7各学历的招聘人数总和柱状图

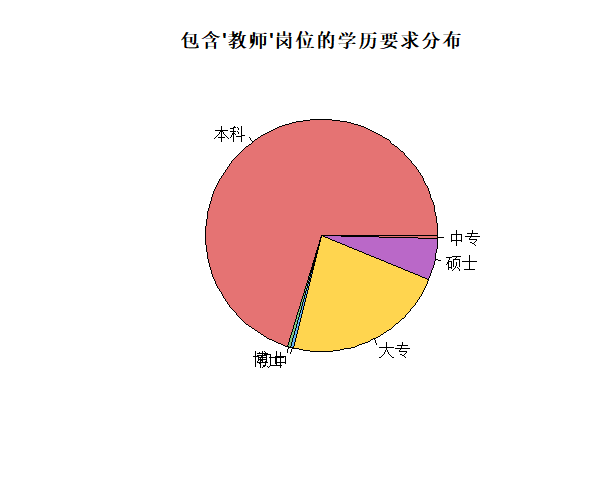


该图考虑到了不同招聘信息的招聘人数不同,反映了各个学历的招聘信息的招聘人数总和,从图中可以看出,从本科提升到硕士和本科学历,能够多接触到的工作机会增加并不显著

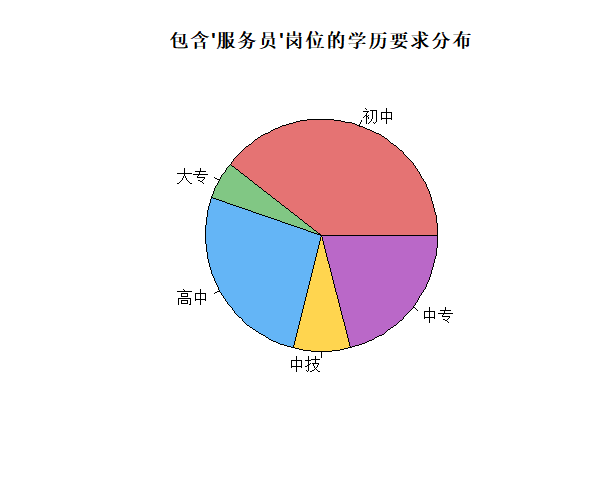
2. 职位分析

这部分可以帮助求职者了解特定职位的需求情况。

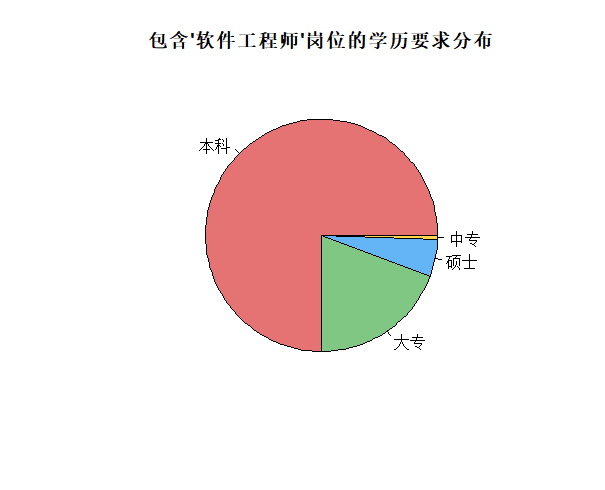
2.1岗位需求与学历关系分析 (以律师,软件工程师,教师,服务员为例)

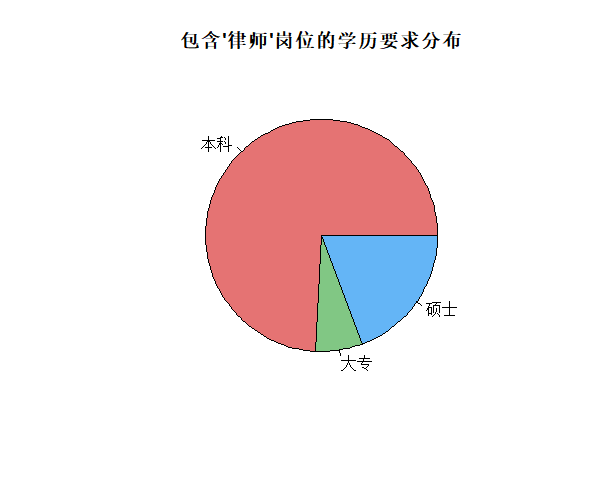


对于教师岗位来说,对学历的要求较高,大部分都要求要本科以上



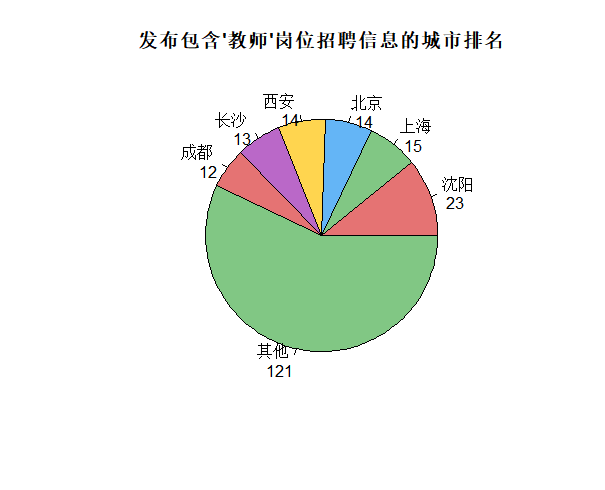
对于服务员岗位来说,学历要求并不高,最高只要求大专学历

对于软件工程师岗位来说,对于学历的要求较高

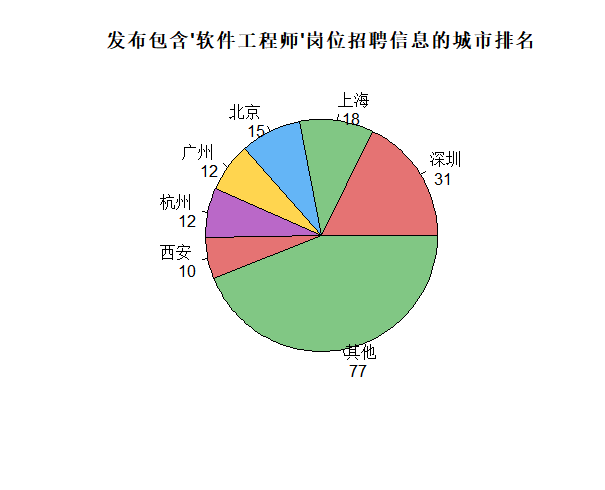
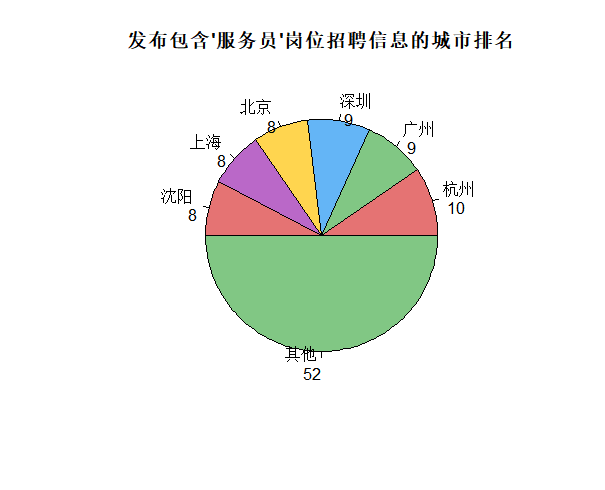
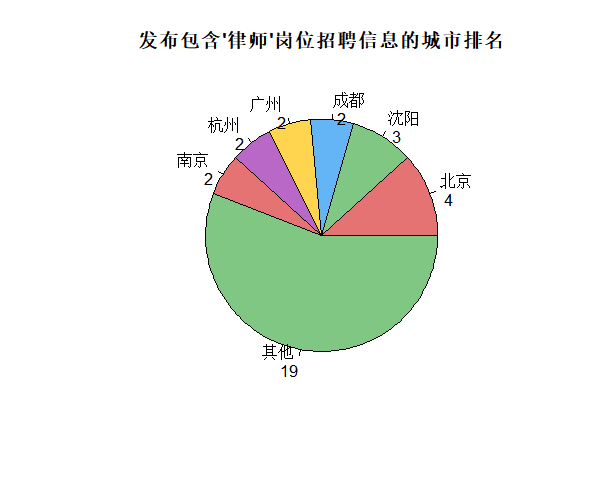


对于律师来说,对于学历的要求很高,很多招聘信息要求至少要本科或者硕士学历

2.2不同城市的岗位需求分析 (以律师,软件工程师,教师,服务员为例)



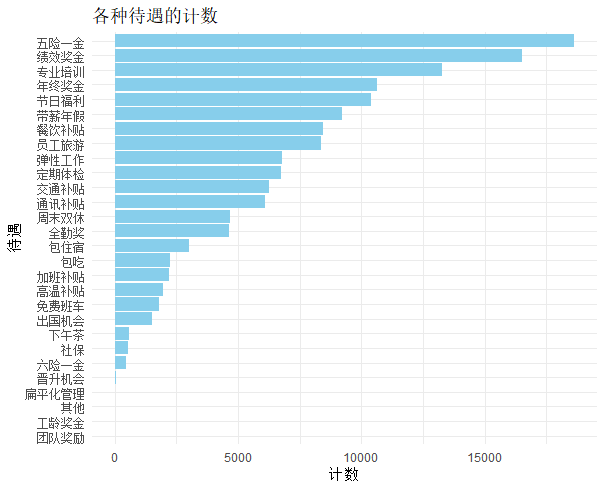
该图反映了包含教师这一岗位的招聘信息数量的城市的排名。从图中可以看出，对于教师岗位的求职者来说,沈阳的招聘信息数量最多，另外，前六名的城市已经约占据了一半以上，说明其他城市的关于教师的招聘信息很少。

类似的，对于软件工程师这个岗位来说，深圳发布的招聘信息数量最多

3. 待遇与福利分析

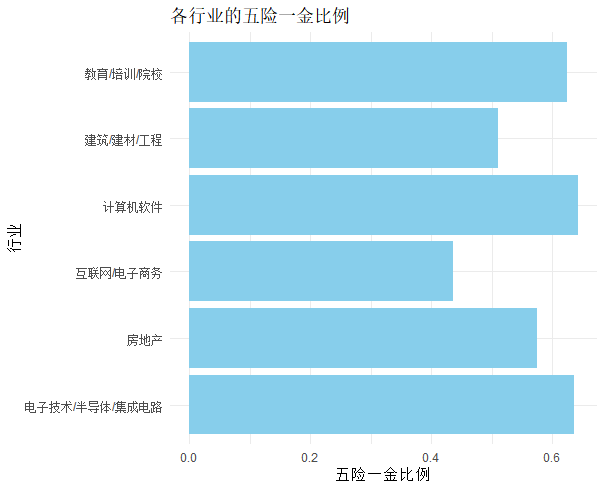
对于求职者，这部分的数据可以帮助他们评估自己的待遇。

3.1常见公司待遇分析



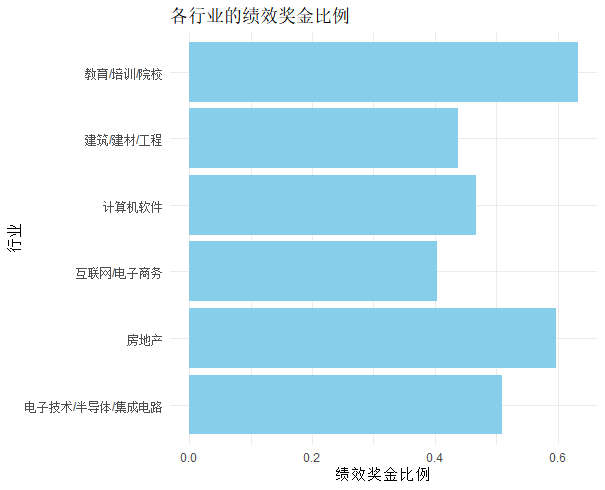
在所有招聘信息提到的待遇中，出现数量最多的依次是“五险一金”，“绩效奖金”，“专业培训”等。福利对于员工来说的重要性有时不亚于薪酬，因此，建议求职者不能一味关注薪酬，求职者应该确保雇主为他们提供五险一金，这设计员工的长期福利。另外，对于需要激励才能够更好发挥工作能力的人，也应该追求具有绩效将近，年终奖金的工作。

3.2不同行业五险一金的比例



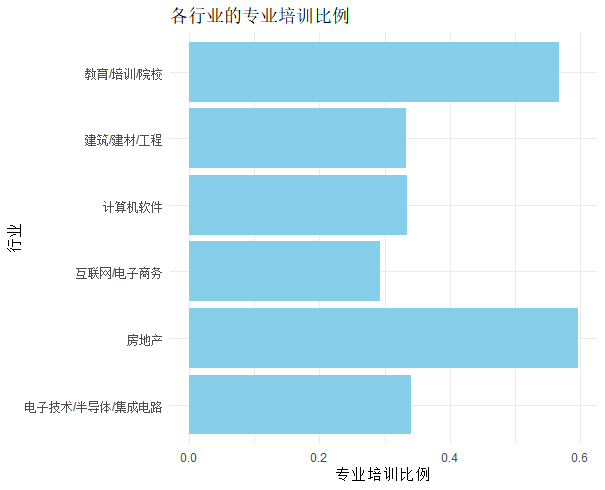
五险一金是最常见的福利之一，但是在前六个热门行业中，含有五险一金的比例并不高，最多只有0.62。建议求职者应该选择五险一金比例较高的行业进行工作，例如计算机，电子技术等。也建议求职者发展这些行业相关的能力，因为能享受到五险一金的概率较高。

3.3不同行业绩效奖金的比例



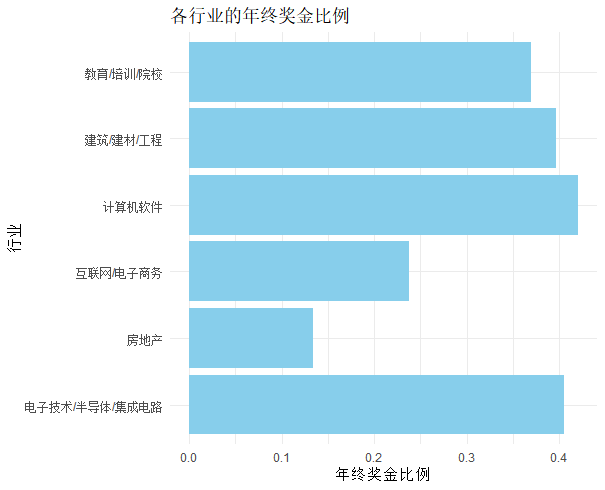
绩效奖金对于需要激励才能更好工作，更好得到满足感的员工来说，选择有绩效奖金这样的待遇的工作，是非常必要的。对于绩效奖金来说，房地产，教育培训等行业的比例较高。

3.4不同行业专业培训的比例



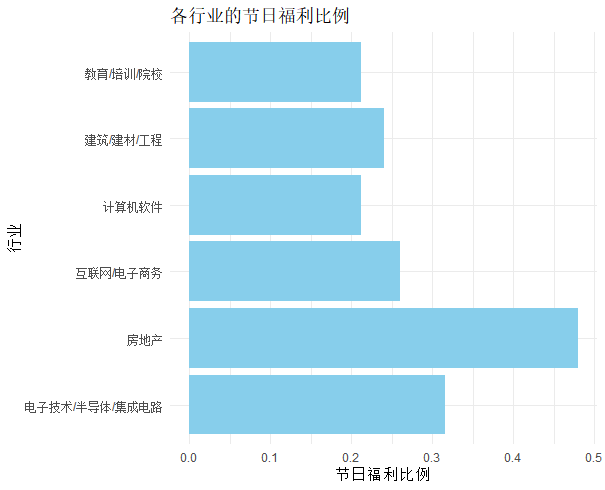
专业培训可以帮助员工提高技能和知识。希望通过工作提升自身能力的求职者应优先考虑。

3.5不同行业年终奖金的比例



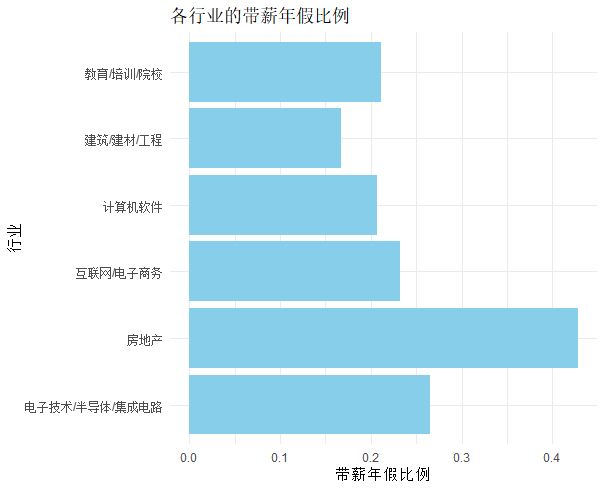
年终奖金是额外的薪酬，重视薪酬的求职者应该关注工作是否有年终奖金。不同行业的年终奖金因为行业原因有所差距。相比之下计算机软件和电子技术的年终奖金比例较高。

3.6不同行业节日福利的比例



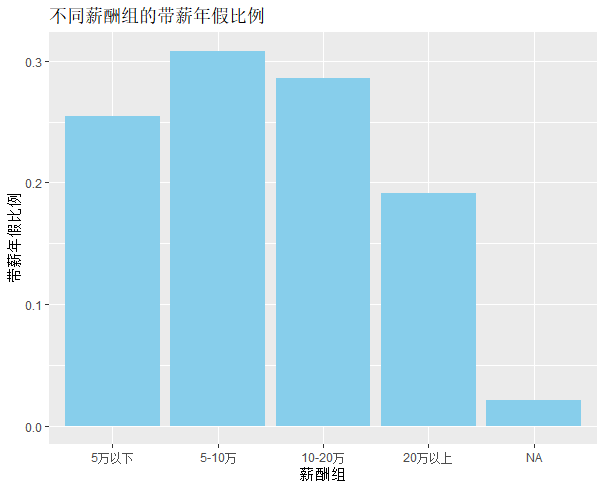
节假日福利反映了公司对于员工的关怀，能够提升员工的满足感，也是福利待遇的一种，房地产的节日福利高于其他行业，而其他五种热门行业的该福利的比例偏低

3.7不同行业带薪年假的比例

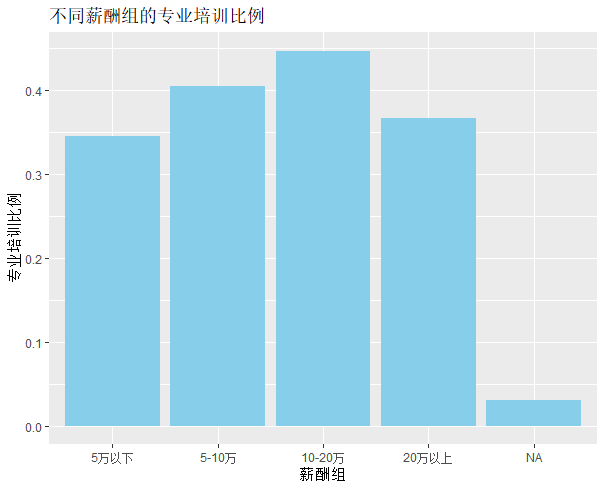


带薪年假也是一种额外薪酬，重视薪酬的求职者应该关注工作是否有年终奖金。相比之下房地产行业的带薪年假比例较高。

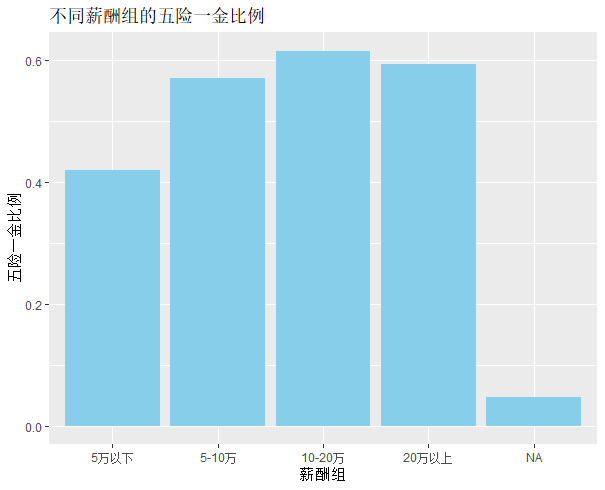
3.8高薪酬是否意味着更好的待遇？用五险一金,专业培训,带薪年假为例



对于带薪年假这种福利来说，薪酬在20万以上的工作带薪年假的比例较低，因此高薪酬并不一定意味着更好的额外薪酬奖励。相比之下，年薪在5-10的工作，带薪年假的比例最高。



对于专业培训这种福利来说，随着薪酬增加，专业培训的比例升高，但是20万以上的工作，专业培训的比例较低

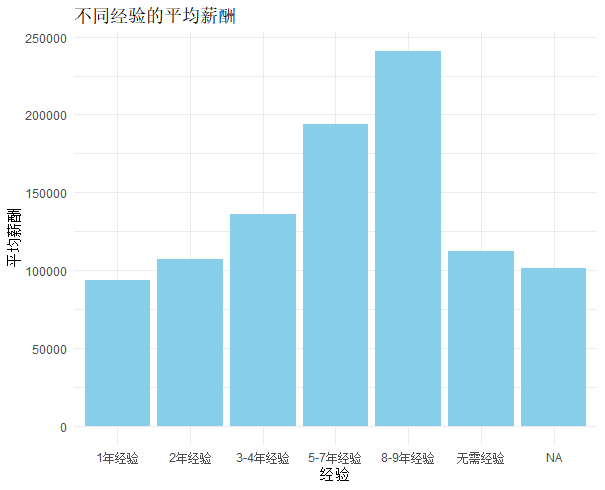


对于专业培训这种福利来说，随着薪酬增加，五险一金的比例升高，因此，高薪酬往往意味着更有可能得到五险一金。

4. 经验与薪酬关系

这部分可以帮助求职者根据自己的经验评估期望的薪酬。

4.1经验对薪酬的影响



从图中可以看出，随着经验的增加，平均薪资有明显的增加趋势。因此，多积累经验对于求职者提升薪酬来说是非常重要的一种方式。

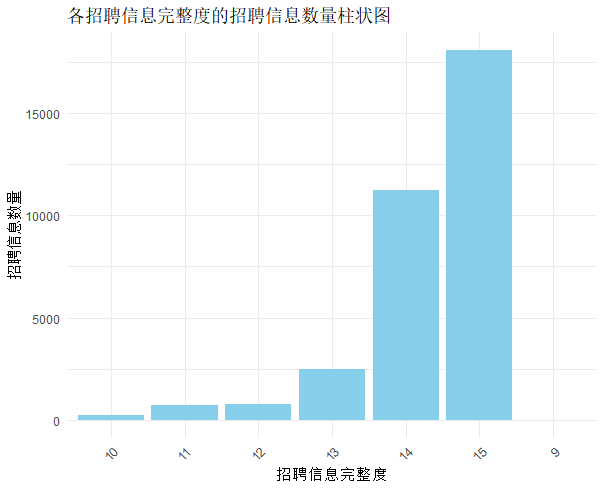
5. 招聘信息完整度选择

首先要定义招聘信息完整度这个概念。

在数据集中，定义"职位描述", "公司名", "公司类型", "规模", "薪酬", "行业", "待遇", "联系方式", "薪酬"几个字段较为重要，是原数据集中就存在的，应占据更多的权重，定义一条数据的这些字段，每有一个不为空，则信息完整度增加2。

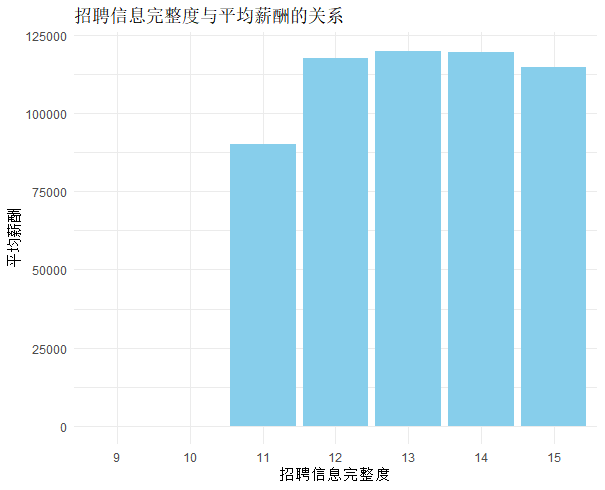
而"经验", "在校生或应届生", "城市", "地区","招聘人数","发布时间","学历"是由地点及需求列拆分得到的，因此应占据较小的权重。定义一条数据的这些字段，每有一个不为空，则信息完整度增加1。

5.1招聘信息完整度分析



大部分招聘信息的完整度都集中在了14和15，意味着大部分招聘下班信息都有一定的不完善。

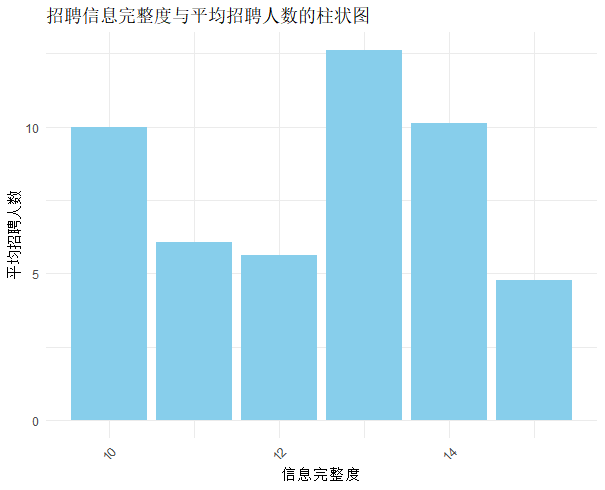
5.2招聘信息完整度与薪酬的关系



信息完整度越高就意味着更高的要求，从而意味着更高的薪资？还是信息完整度越高意味着公司对于员工的需求更为重视，从而意味着更高的薪资？

从结果来看，并不是信息越完善，薪酬就越高，信息完整度在12-14之间时，平均薪酬最高。

5.3招聘信息完整度与招聘人数的关系

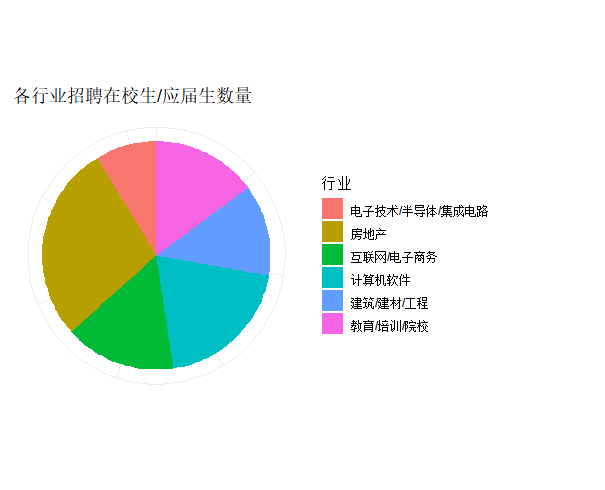


是否信息完整度越高，是因为要招聘的人数很多，所以需要讲得更加清楚？结果发现信息完整度与招聘人数并不存在很强的关系

6. 行业与城市选择

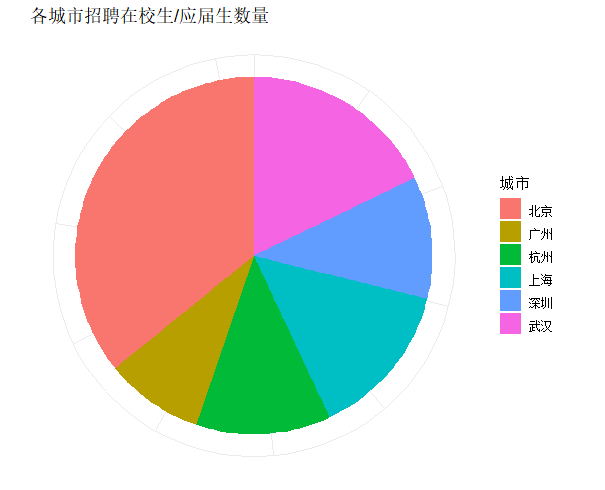
这部分可以帮助求职者选择合适的行业和城市。

6.1哪些行业为在校生和应届生提供更多的机会



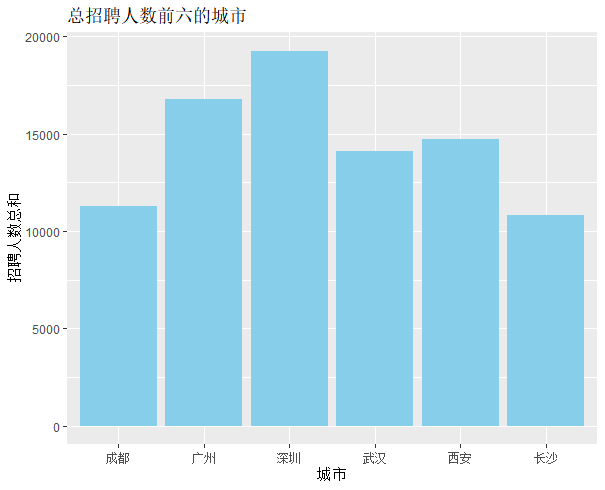
对于在校生和应届生来说，他们往往没有工作的经历，对于他们来说，他们可以选择这种专门写着接收在校生/应届生的招聘信息。而发布这些信息最多的行业是房地产，其次是计算机软件和互联网电子商务。

6.2哪些城市为在校生和应届生提供更多的机会



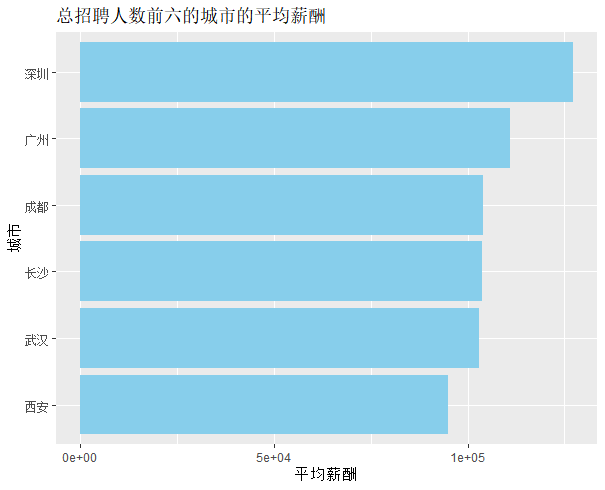
对于在校生和应届生来说，招聘最多在校生/应届生的城市是北京，占据了很大一部分,且约为第二名的两倍,其次是武汉，深圳。

6.3总招聘人数前六的城市



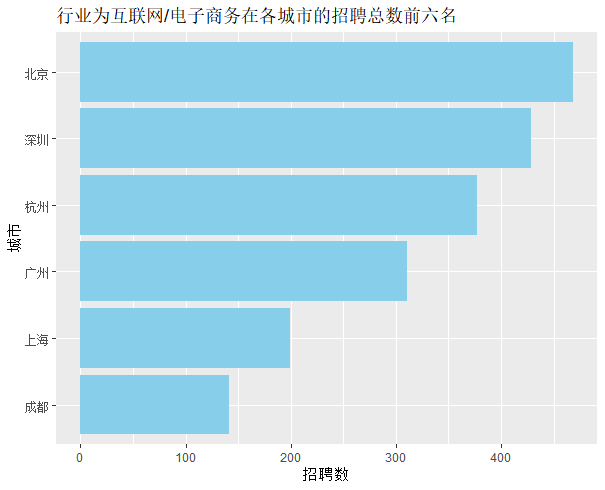
总招聘人数考虑了不同招聘信息的招聘人数不同,该分析展示了各个城市的招聘总人数的对比信息。其中深圳的招聘人数最多，这意味着更多的就业机会，相比于招聘信息数量，更能够反应就业的机会。因此这对于想要在异地求职的求职者的城市的选择具有很大的帮助。

6.4总招聘人数前六的城市的平均薪酬

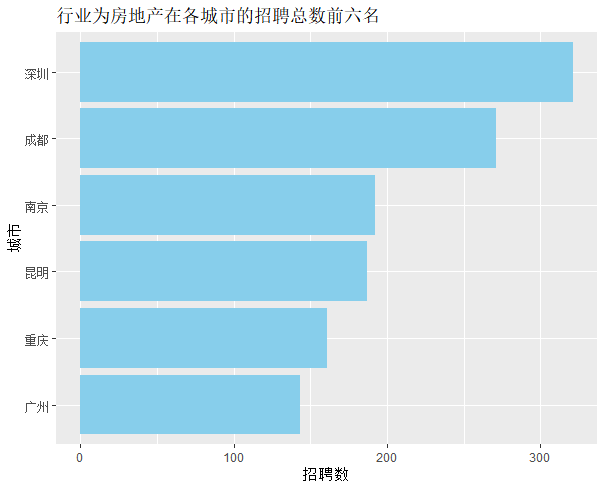


总招聘人数平均薪酬考虑了每个招聘名额的平均薪酬，而不是每个招聘信息的平均薪酬。因此对于求职者预测城市的薪酬具有更加有效的作用。从分析来看，深证，广州的平均薪酬最高。

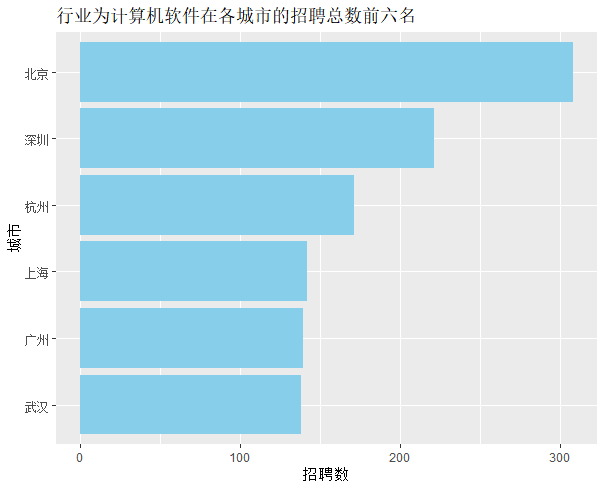
6.5前六行业与在各城市的招聘总数的前六名

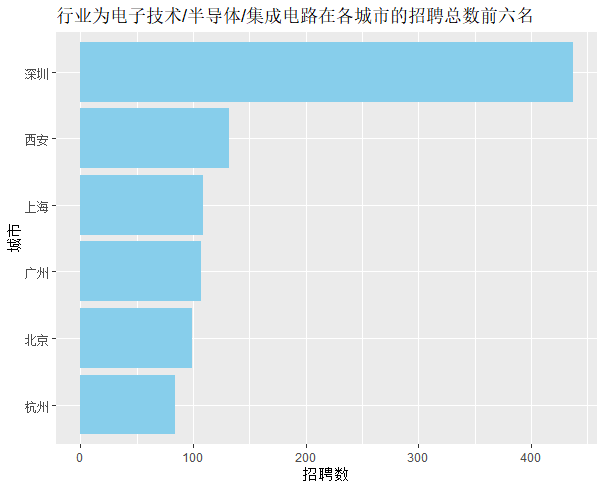


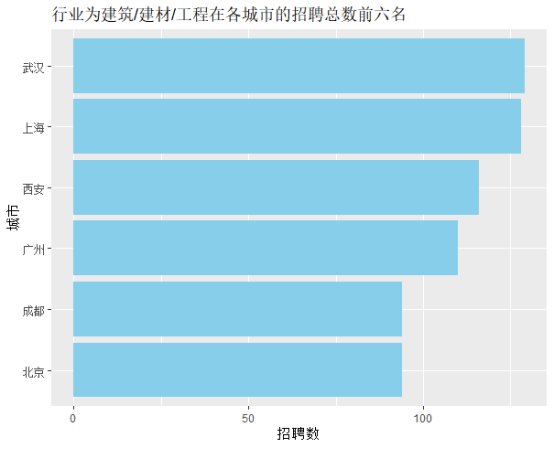
这反映了互联网/电子商务相关行业知识的求职者,可以选择这六个城市进行应聘,因为招聘总数更多。



同样地，这个分析能够为房地产行业的求职者分析再该行业，哪些城市的工作机会更多

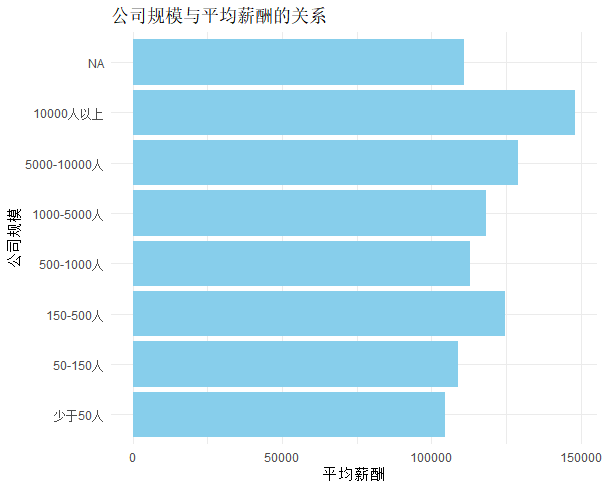






7. 公司规模的选择

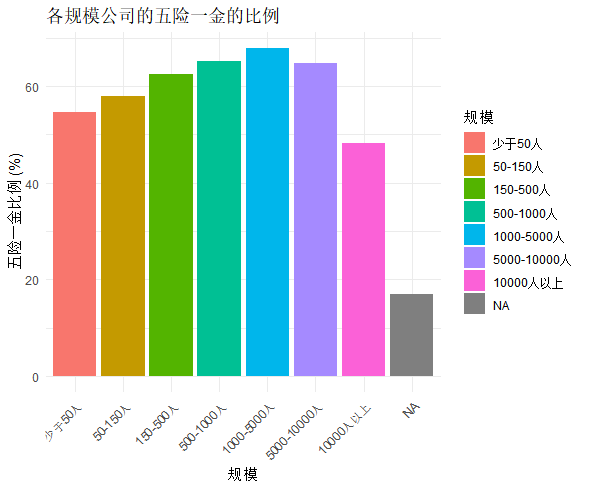
7.1公司规模与薪酬的关系



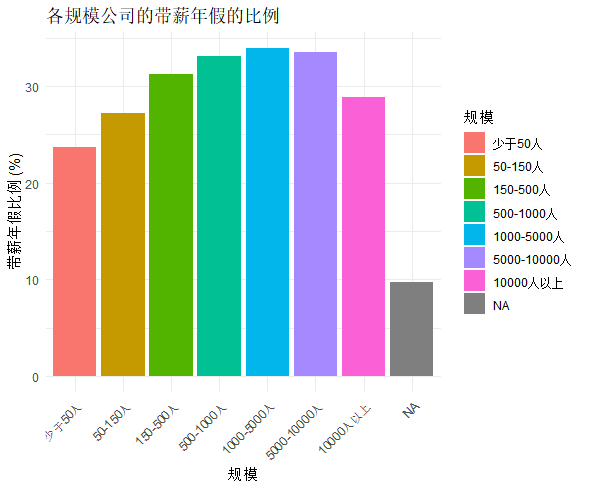
对于求职者来说，公司的规模也是需要考虑的因素。从数据显示，公司规模与薪酬具有较强的关联性，公司规模越大，平均薪酬就相对更多。

7.2公司规模与待遇的关系

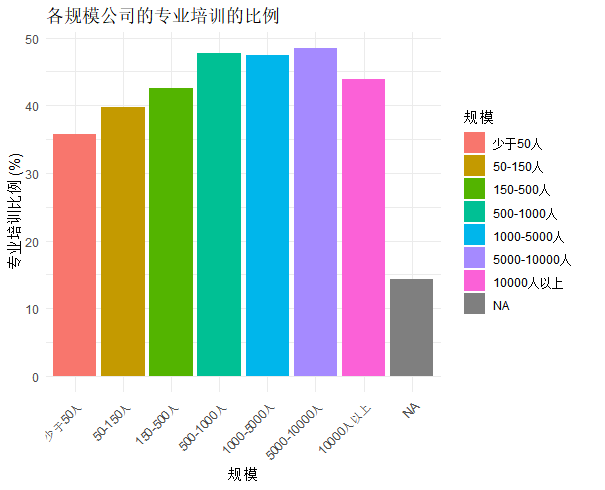
求职者可能有一种刻板印象，就是公司越大，就越正规，那么相应的待遇也就越好，于是通过数据分析，验证是否是这样，并给予结果给出求职者相应的建议。



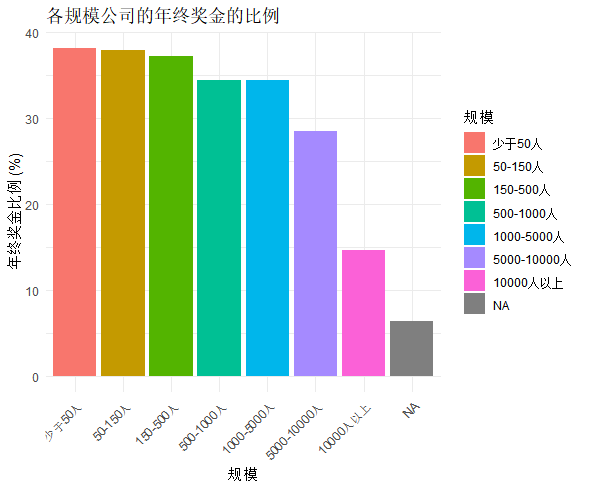
对于五险一金来说，确实是公司规模越大，五险一金的比例就越高，但是10000人以上大公司不符合这个规律，反而是五险一金率最低的。



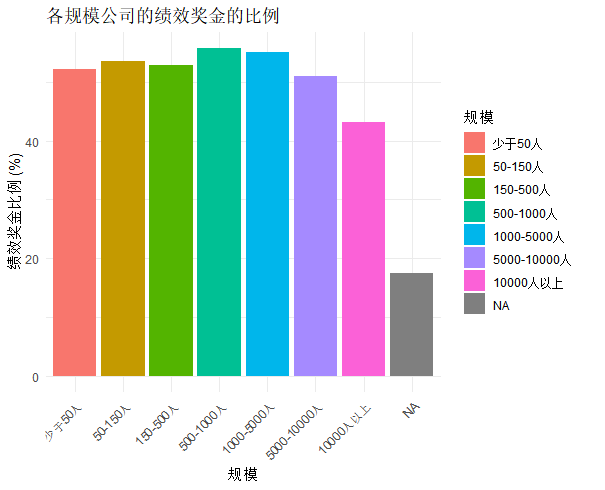
对于带薪年假来说，也符合公司规模越大，带薪年假的比例就越高，同样10000人以上大公司不符合这个规律，其带薪年假的比例处于50-150人的规模和150-500人的规模之间。



对于专业培训来说，也符合公司规模越大，专业培训的比例就越高，同样10000人以上大公司不符合这个规律



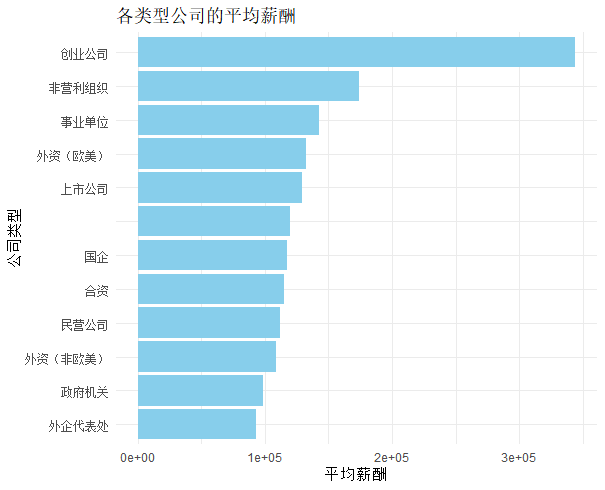
对于年终奖金这个待遇来说，不符合公司规模越大，年终奖金的比例就越高的规律，反而呈现明显的递减规律。



对于绩效奖金这个待遇来说，其比例与公司规模没有很强的关系，各个规模的公司的绩效奖金的比例大致相同，10000人以上的大规模公司的绩效奖金比例最低。

8.公司类型的选择

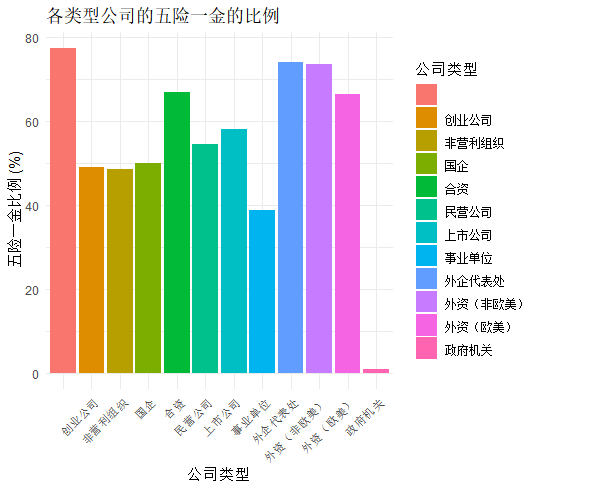
8.1公司类型与薪酬的关系

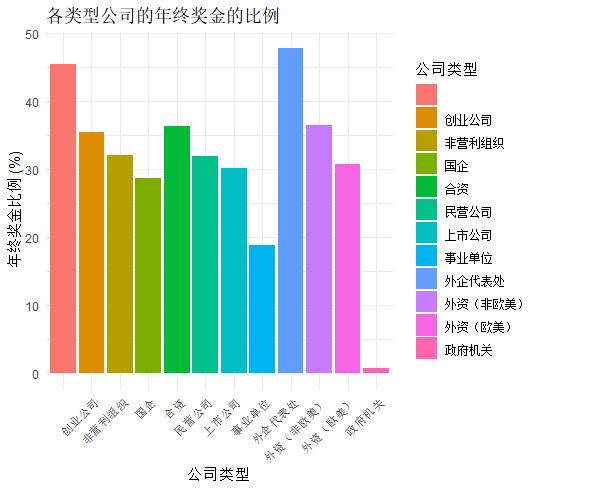
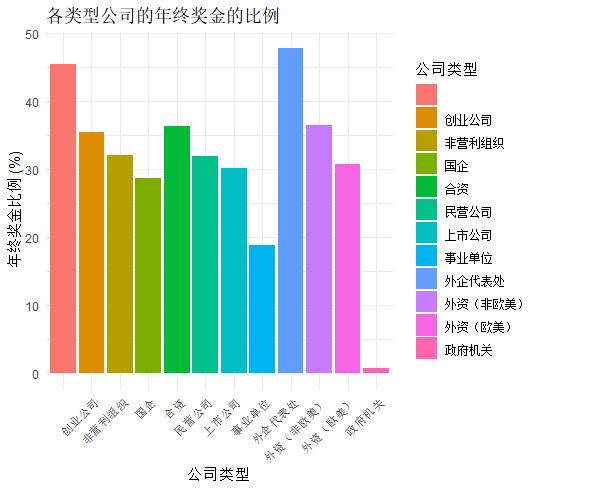
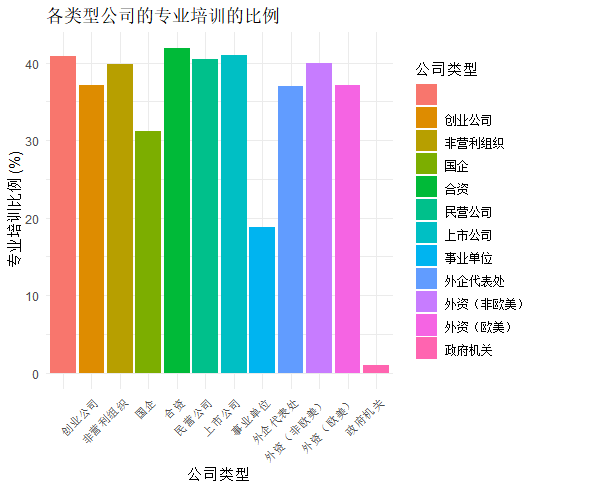
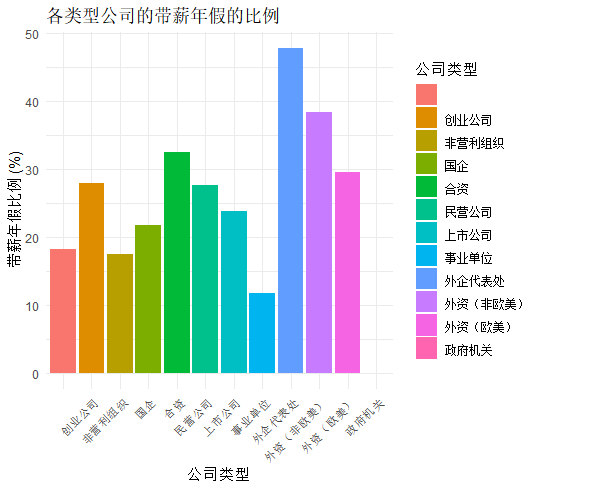


公司有着不同的性质，数据分析显示，公司的性质与薪酬有着很强的关系例如创业公司的平均薪酬远高出其他类型的公司，而常见的国企，民营公司的平均薪酬较低，很多人说外资企业薪酬高，但是数据显示欧美的外资公司薪酬较高，但是非欧美的外资很低甚至还低于国企和民营公司。而很多人看到这句话都不会区分欧美和非欧美的外资。因此这句话具有误导性，并不完全正确。对求职者的建议是，如果想追求更高的薪资，可以选择创业公司或者欧美的外资企业。而政府机关的薪酬很低。

8.2公司类型与待遇的关系

都说在国企工作更加稳定，但是国企能否给予比其他企业更好的待遇？





从中可以看出，国企的各种待遇比例都比较低，同样常见的民营公司和上市公司的各种待遇的比例都略高于国企。因此，国企的待遇并非很高。通过这一部分的分析，可以从待遇这个角度出发，帮助求职者进行公司类型的选择。