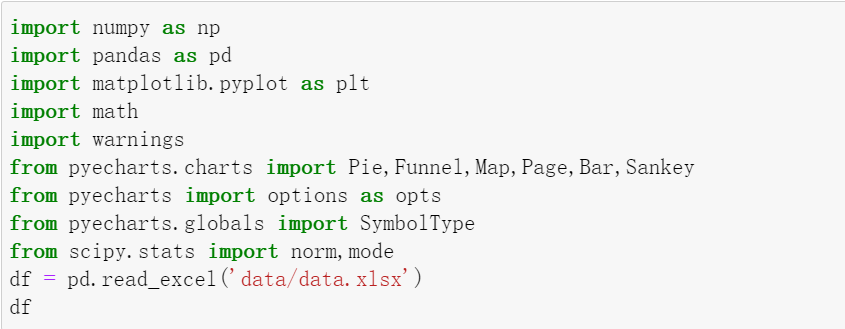
招聘岗位分析

1. 数据采集

分析所用的数据由网上获取。

代码：



导入的数据表：



数据的属性一共有type，name，company，salary，city，experience，edu组成，分别对应着招聘岗位的工作类型，工作名称，所在公司，薪资，城市，工作经验和学历。

二、数据清洗

导入的数据并不一定是理想的数据，可能带有重复的数据或者是缺失的数据，所以需要进行数据清洗，去除重复的数据和带有空值的数据，防止干扰到之后的分析。

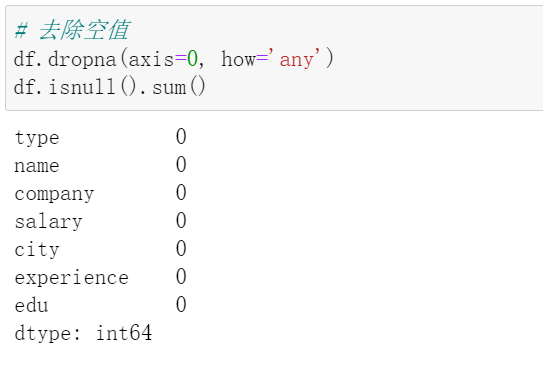
1. 去除重复的数据

代码：



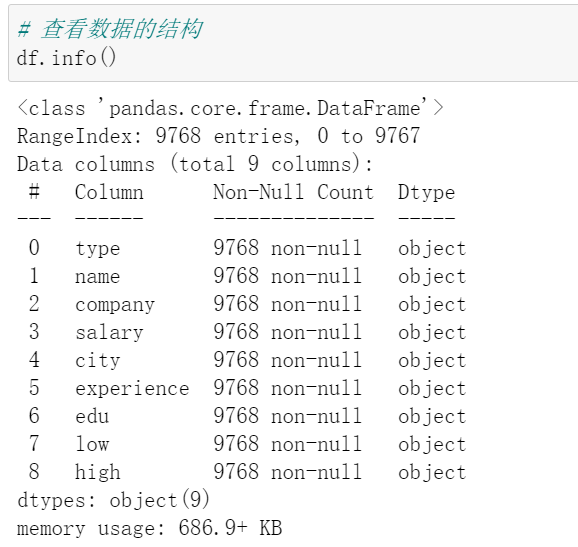
1. 去除带有空值的数据

代码：



1. 查看数据的结构

代码：



数据清洗完成，具备进行分析的条件。

三、数据建模

数据中的薪资表示的是一个范围，所以为了方便分析，需要计算出薪资的平均值。

1. 计算薪资范围的最低值和最高值

这里采用pandas中的str.extract()函数截取需要的数字。

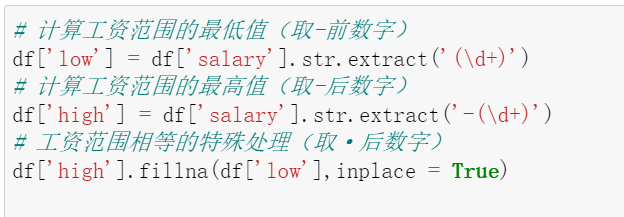
正则表的式str.extract('(\d+)')中的()表示需要截取的内容的位置，\d+表示截取多个数字。

这里(\d+)表示的是截取从一开始到后面非数字的全部数字，例如数据20-40k，此处取出的就是20。

这里-(\d+)表示的是取出范围符号‘ - ’后的全部数字，例如数据20-40k，此处取出的就是40。

如果薪资为固定的数字，例如20，则-(\d+)则取不到任何值，因为没有范围符号‘ - ’，所以直接用低值填充，这样将来计算平均值的时候还是原来本身。

代码：



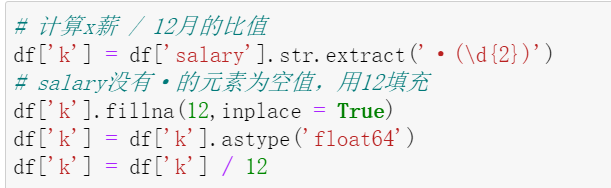
(二)计算每个月发薪资次数的平均值

并不是所有薪资都是每个月发一次的，例如20-40K·17薪表示一年发17次工资，所以需要计算平均每个月发k次工资，其中k为x薪与12月的比值。

计算步骤是先利用函数str.extract()取出每年发x次工资，之后全部除12计算出每个月发薪资次数的平均值。

str.extract(·(\d{2}))表示取·后的全部数字，例如20-40K·17薪，取出的为结果17。如果数据是20-40K则默认每个月发一次工资， 但是缺少符号·将导致最终计算的结果为空值，所以全部用12填充。由于取出的数据全部是字符串，并且是先计算一年发工资的次数，所以最终需要转换为数字类型再除12计算出每个月发薪资的次数的平均值。

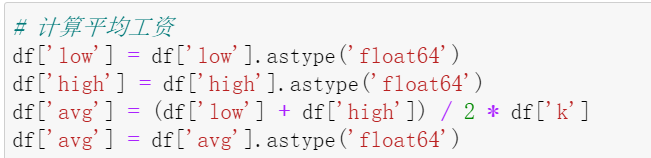
代码：



（三）计算平均薪资

计算公式为（最低值 + 最高值）/ 2 \* （x薪 / 12）

代码：



结果：

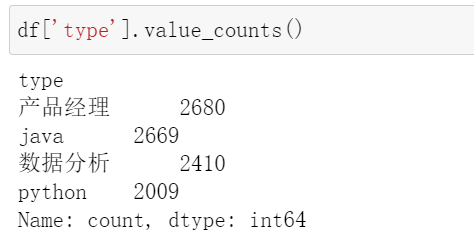


薪资的平均值计算完成，可以进行数据可视化分析。

四、数据可视化

(一) 岗位类型与薪资

查看岗位类型代码：

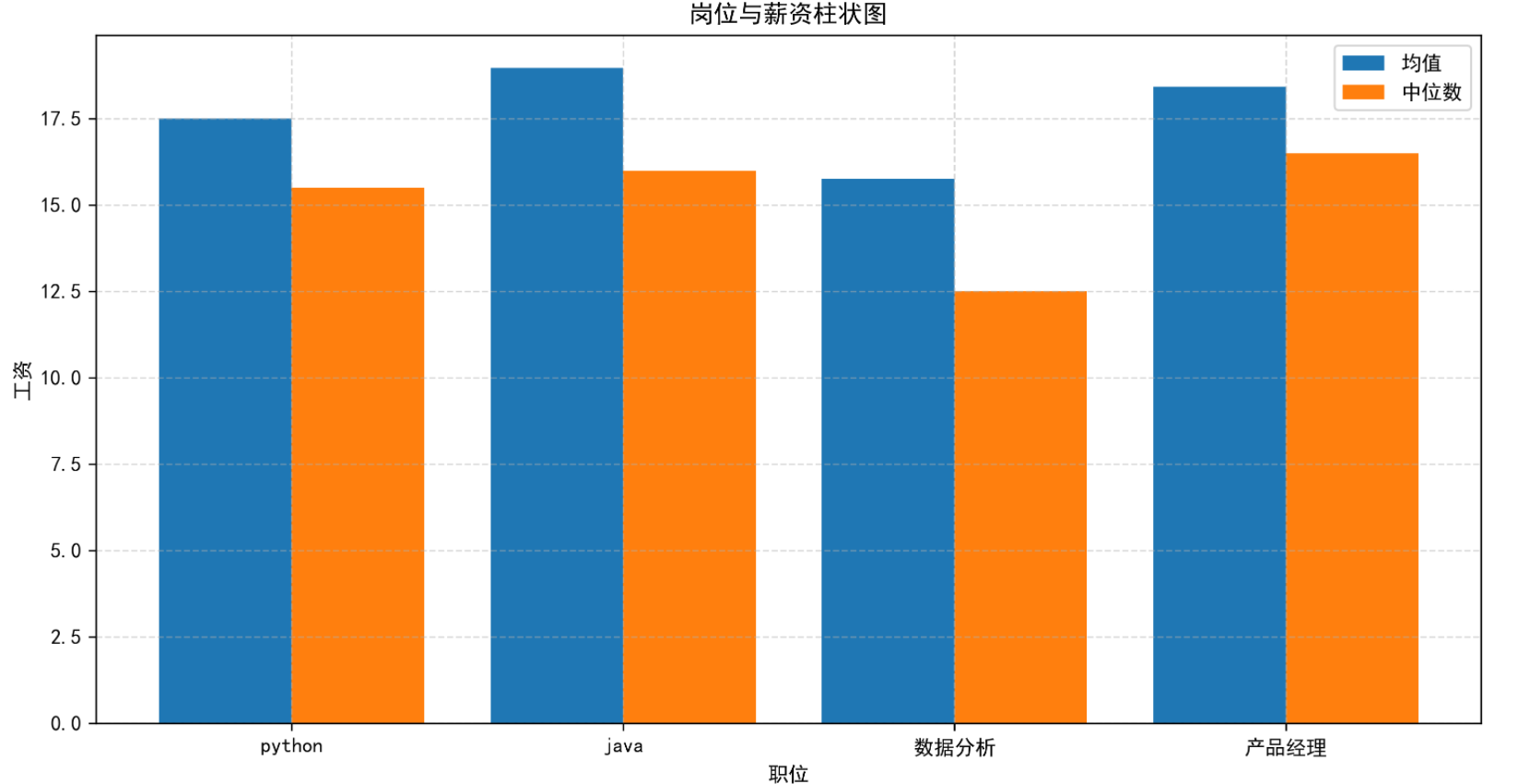


数据一共包括4种不同类型的岗位，分别为产品经理、java、数据分析和python。

岗位类型与薪资柱状图代码：



岗位类型与薪资柱状图：



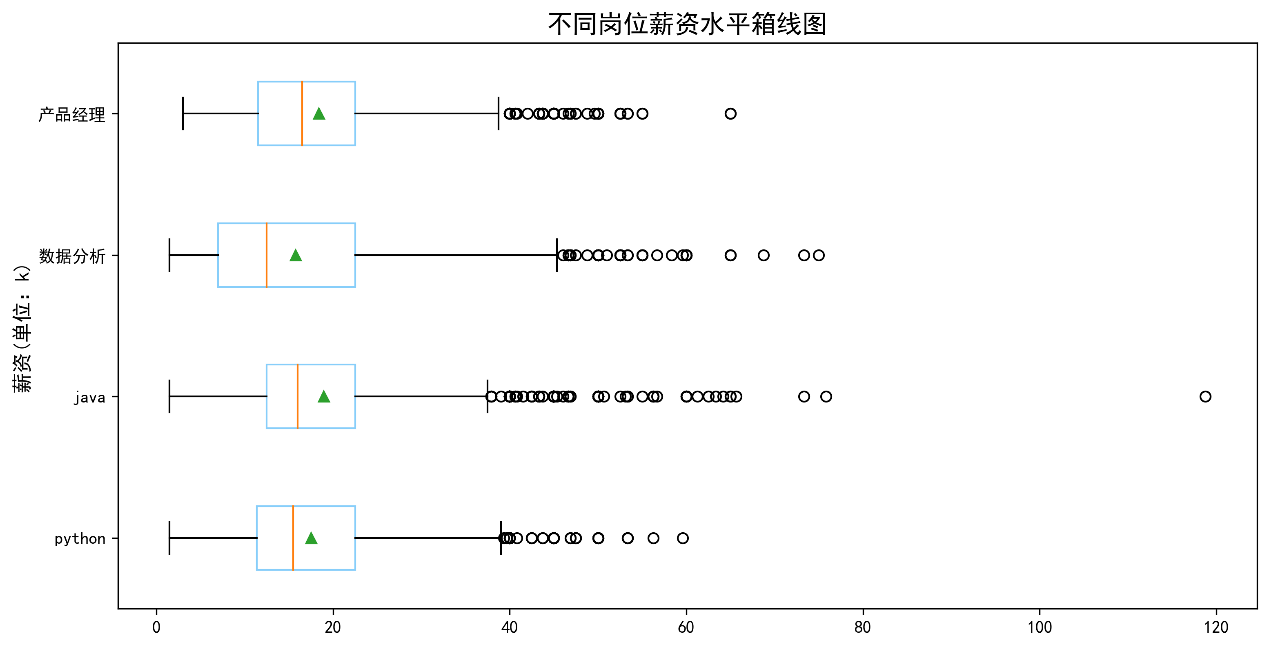
分析：

四个岗位的薪资均值都在15k-19k，最高的是java，约为18.5k左右，最低的是数据分析，约为15.5k左右。Python、java、产品经理的薪资中位数大致相同，约为1.6k，但是数据分析的薪资中位数比前三个略低2.5k左右。数据分析的整体薪资略低2k。

岗位类型与薪资箱线图代码：



岗位类型与薪资箱线图：



分析：

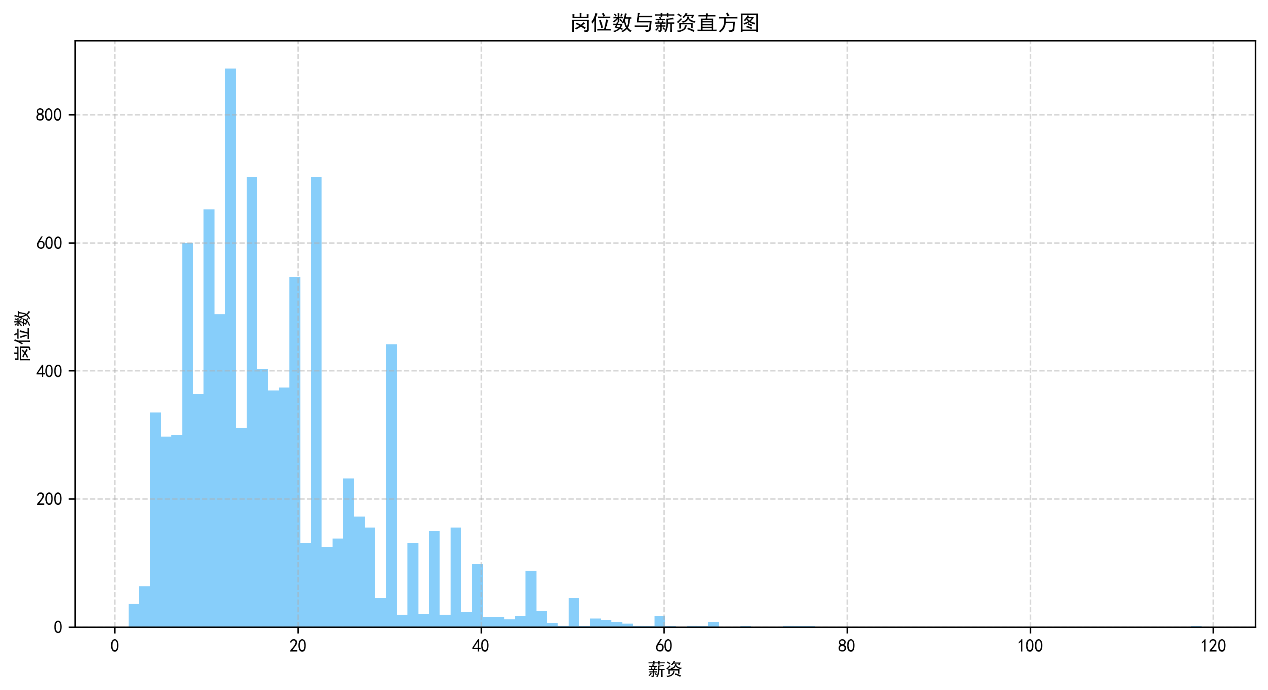
产品经理、java、python的薪资分布较为均匀，数据分析的薪资两级分化较为严重，拿高薪的甚至人数要多于其他三个，但是整体水平较低。同时四种职位的异常点都比较多，说明只有少部分人可以拿非常高的薪资。

(二) 岗位数与薪资

岗位类型与薪资直方图代码：



岗位类型与薪资直方图：

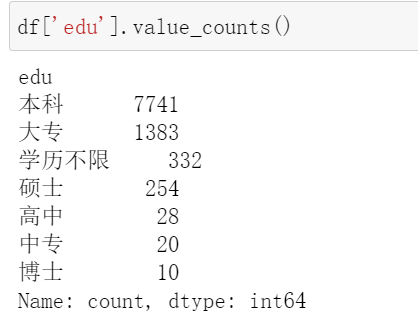


分析：

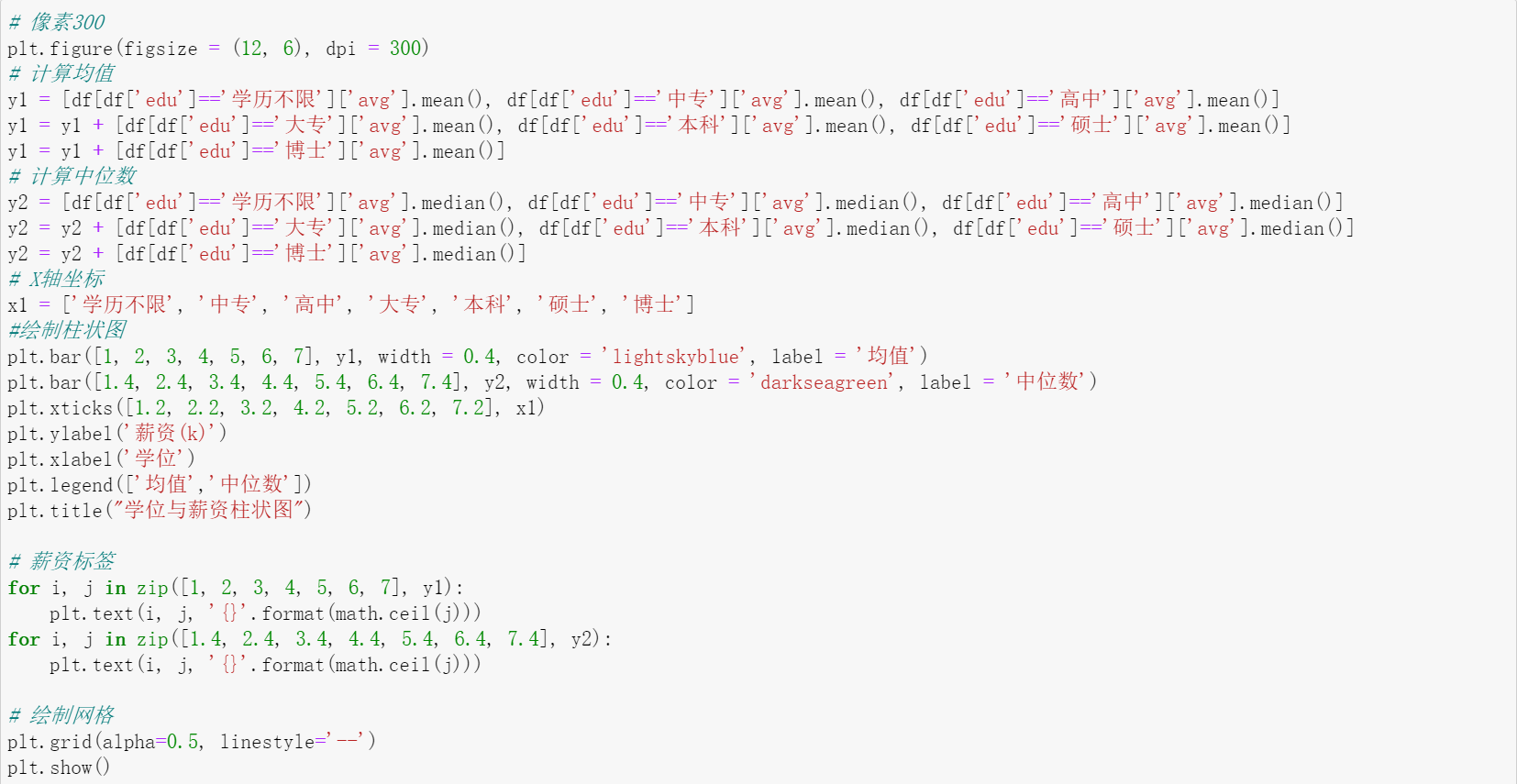
从图中可以得知，绝大多数的岗位薪资都集中在7k-20k之间，再往上就会出现断崖式下降，随着薪资越来越高，其提供的岗位数则越来越少，从几百个到个位数。说明只有极少部分人能拿到非常高的薪水。

(三) 学历与薪资

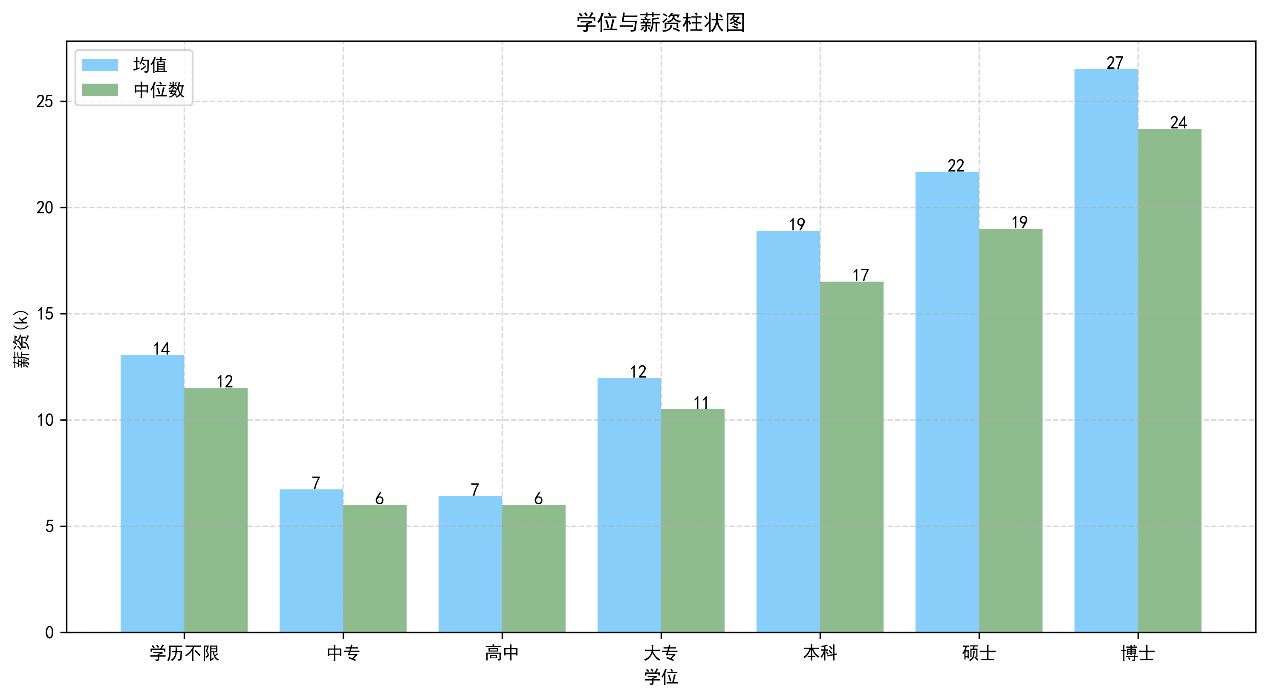
查看学历类型代码：



学历与薪资柱状图代码：



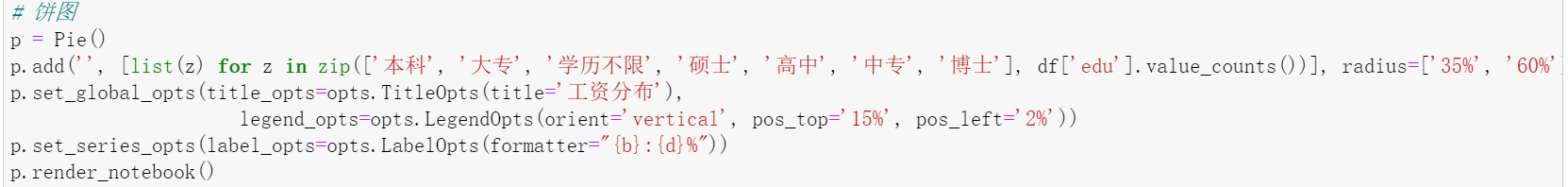
学历与薪资柱状图：



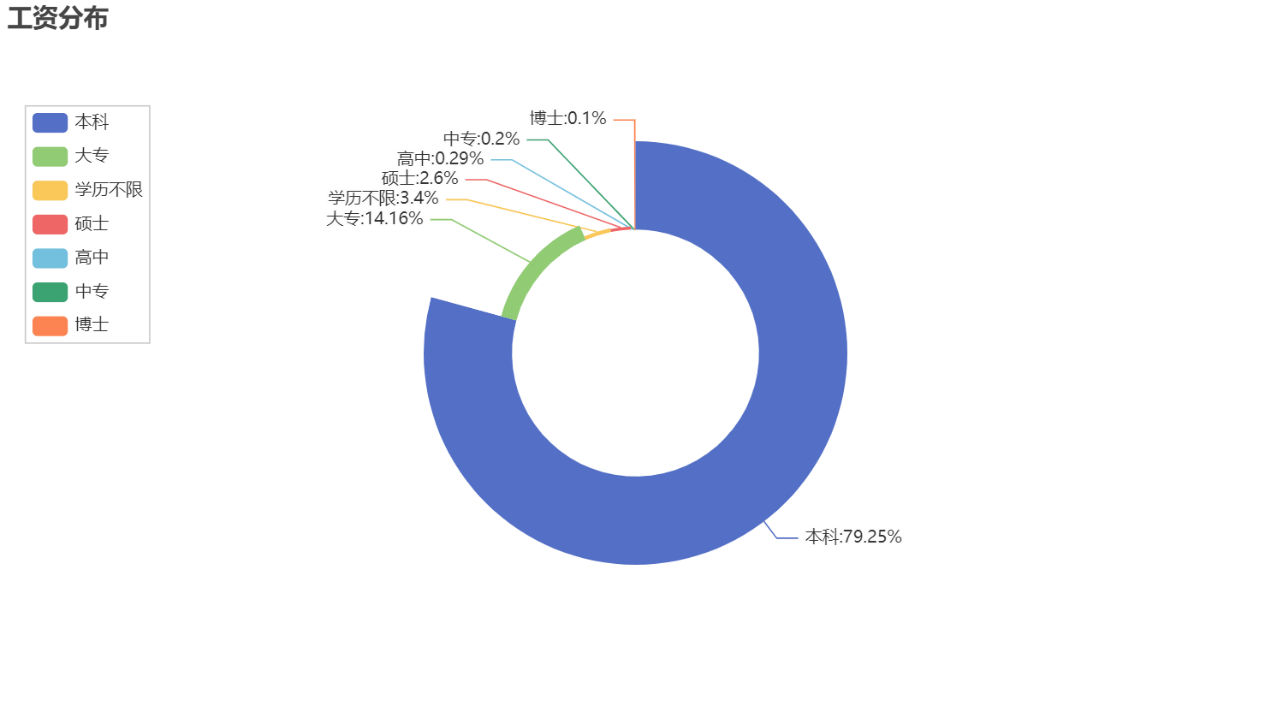
分析：

薪资随着学历的增加而增加，学历不限并不是最低的，略大于大专而低于本科。所以薪资和学历高低呈正相关。

学历饼图代码：



学历饼图：

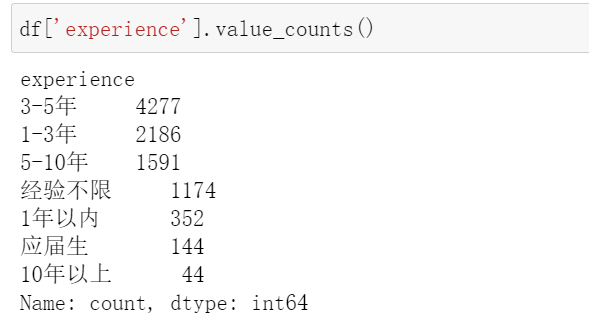


分析：

本科占所有岗位的绝大多数，其次是大专、学历不限、硕士、高中、中专、博士。博士占比仅仅只有0.1%。所以学历取得本科是大多数岗位要求的前提条件。

(四) 工作经验与薪资

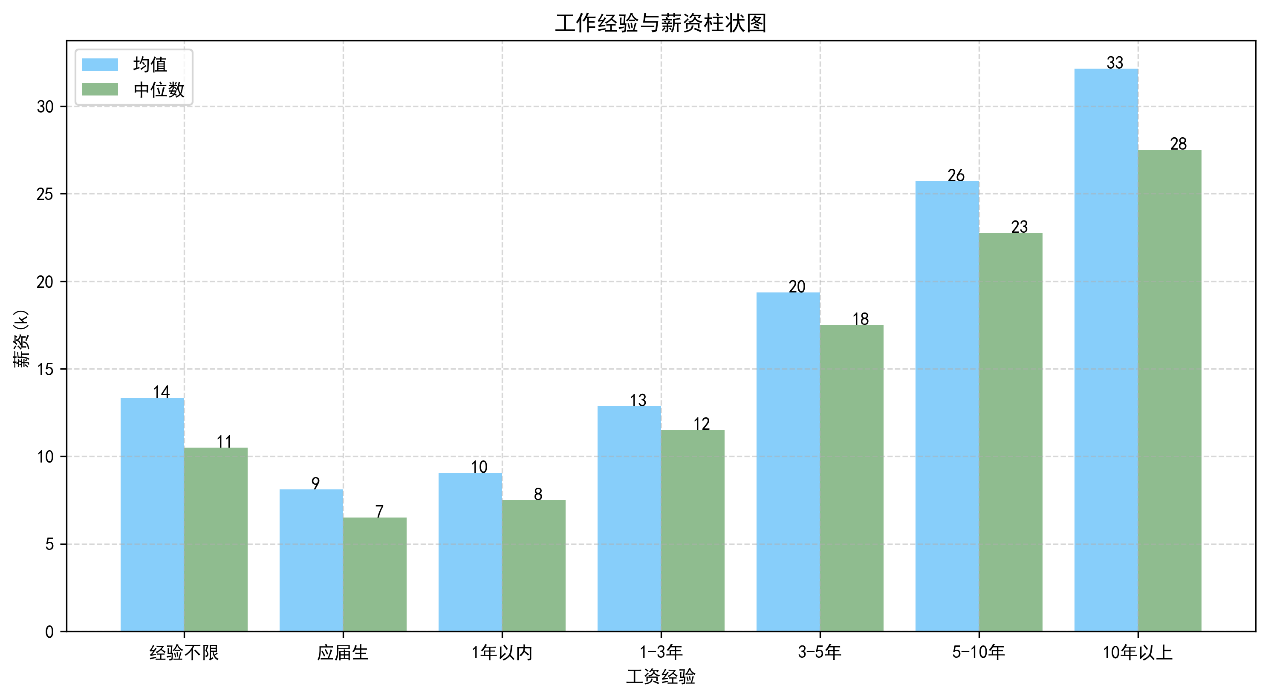
查看工作经验类型代码：



工作经验与薪资柱状图代码：



工作经验与薪资柱状图：

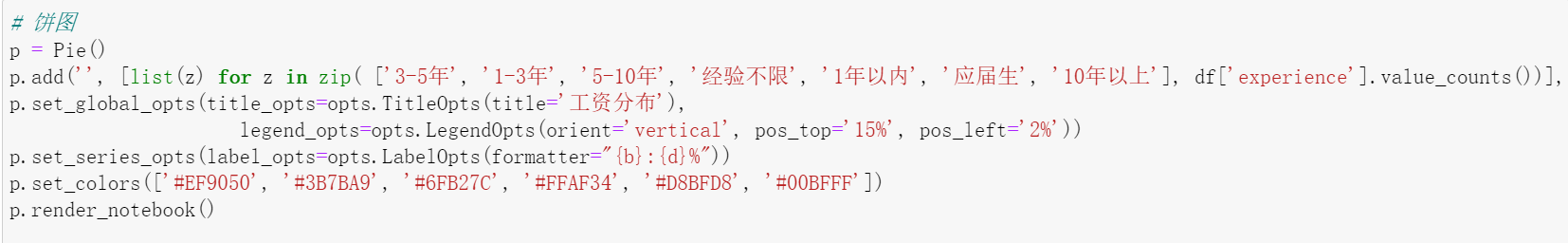


分析：

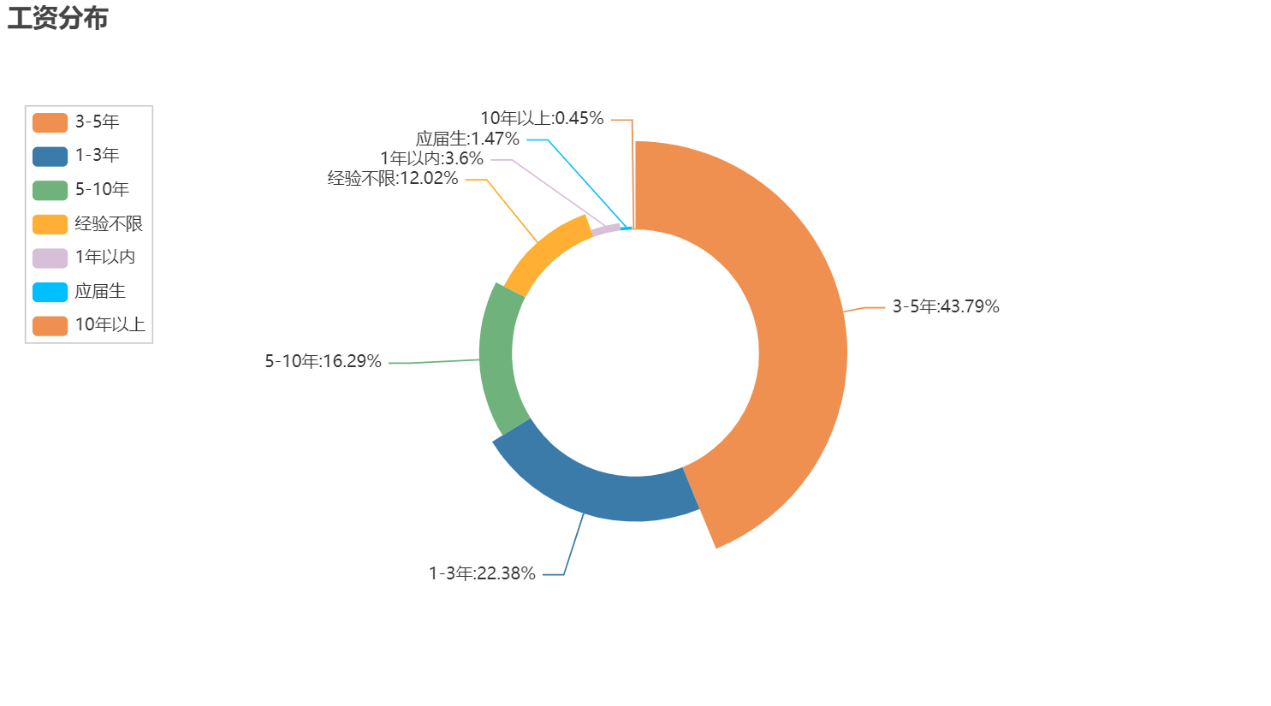
薪资随着工作经验的上升而上升，经验不限和1-3年工资经验大致相同。

工作经验越高，薪资基本上就越高。

工作经验饼图代码：



工作经验饼图：

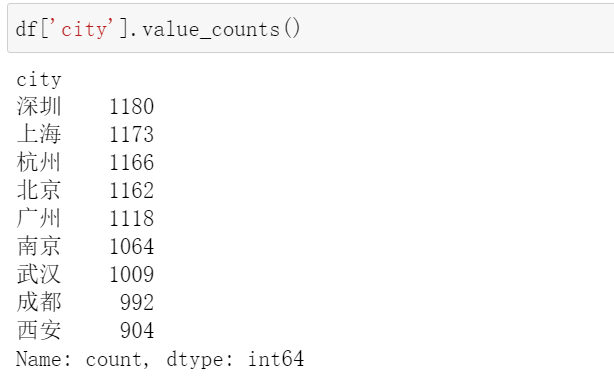


分析：

3-5年工资经验提供的岗位最多，其次是1-3年的工作经验。应届生有且仅有1.47%，经验不限占比12.02%。所以多数岗位对工作经验的要求是3-5年，只有3-5年的工作经验才有资格胜任大部分岗位。

(五) 城市与薪资

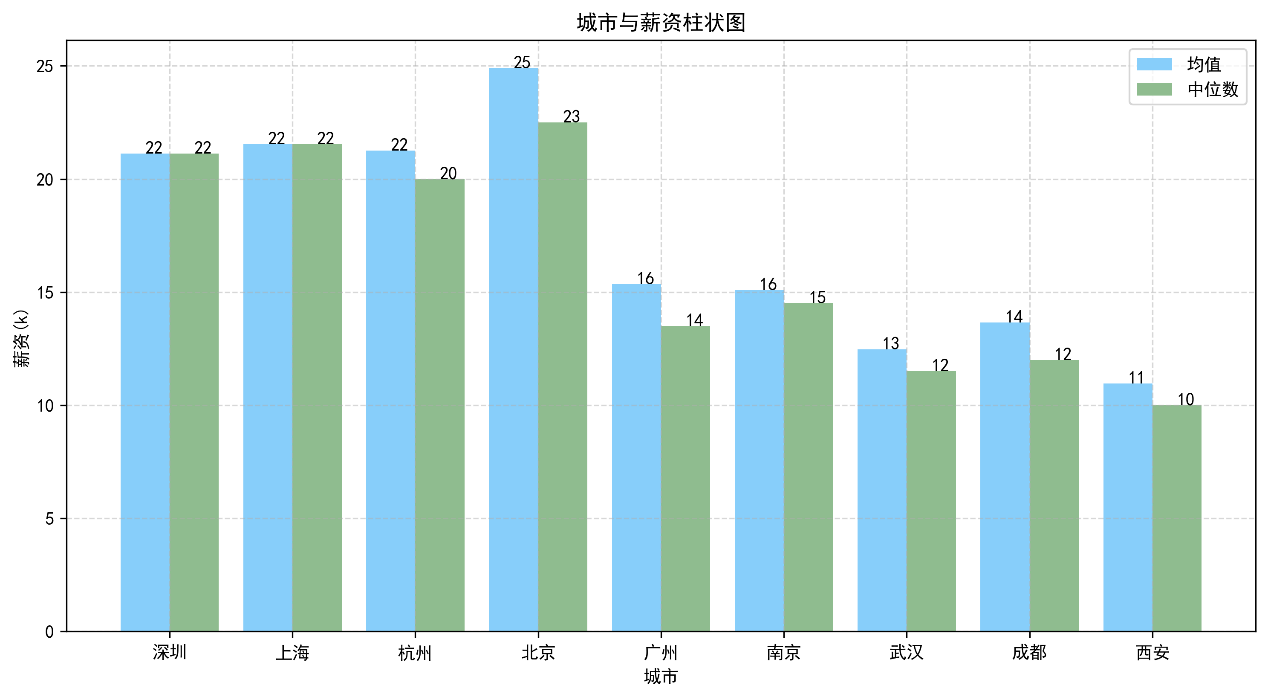
查看城市类型代码：



城市与薪资柱状图代码：



城市与薪资柱状图：



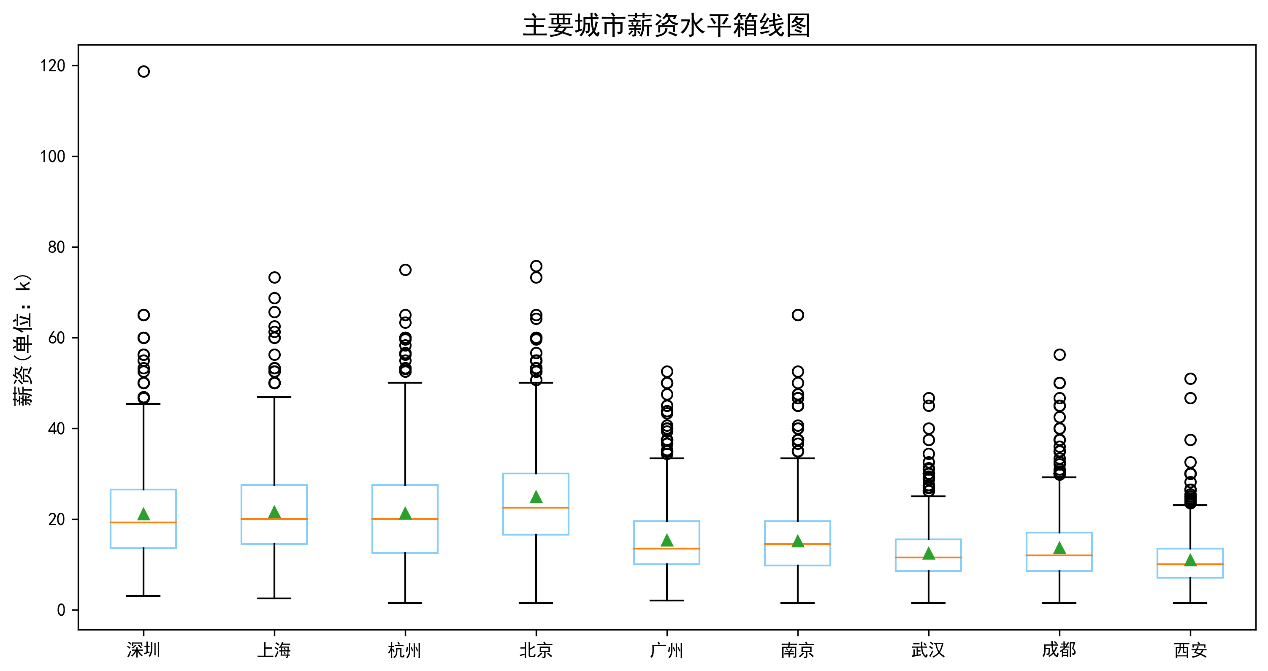
分析：

深圳、上海、杭州、北京要明显高于广州、南京、武汉、程度、西安。说明IT行业的薪资与城市的科技水平相关。城市的科技水平越高，平均薪资就越高。

城市与薪资箱线图代码：



城市与薪资箱线图：



分析：

深圳、上海、杭州、北京的综合薪资要明显高于广州、南京、武汉、成都、西安。但是广州、南京、武汉、成都、西安仍然有人能远超深圳、上海、杭州、北京的高水平，说明个人能力并不局限于城市的科技水平。城市的科技水平只能反映整体的情况。能力强的人同样可以取得非常高的薪资。

五、总结

在python、java、数据分析、产品经理这几个主流的IT行业中，薪资和行业类型的关系不是特别的大，四个行业的薪资平均水平都没有明显高于其他各个行业，但是薪资和学历、工作经验以及城市科技水平有着比较大的关系。大多数人拿的薪资都在10k-20k之内，只有少数人才能拿比较高的薪水。

一个人的学历越高、工作经验越丰富，薪资也就越高。同时城市的科技水平越发达，整体的薪资水平就越高。一个人的工作经验和学历是这个人能力的直接体现，但是科技相对落后的城市又有人能达到甚至超过科技发达城市的高薪资水平。

这证明了薪资很大程度上和一个人的能力是呈正相关的关系。

所以总的来说城市的科技水平决定了薪资的下限，反映了绝大多数人的真实薪资状况；而薪资的上限由一个人的能力决定，能力越强的人拿的薪资就越高，并不会受城市的发达水平约束，在科技相对落后的城市中，能力最强的人的薪资超过发达城市中能力较强的人的情况也是真实存在的。