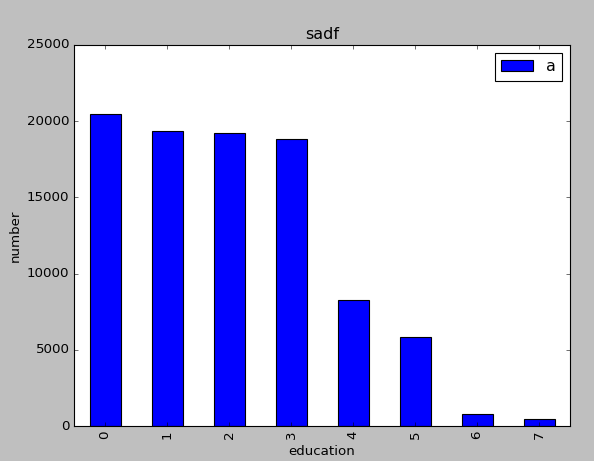
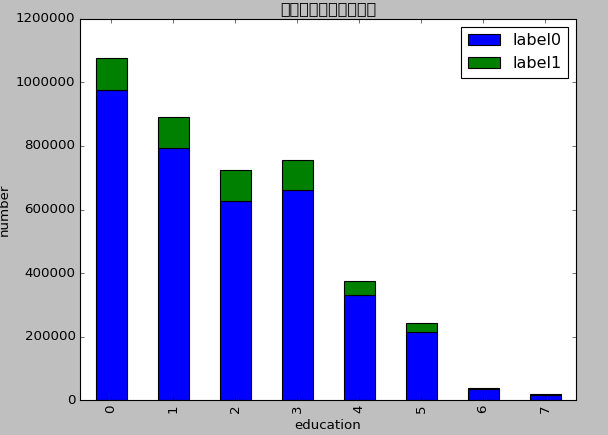
每个特征的第一个图是Label = 1时的分布情况

图二是label的0,1对比，label=1的数量太少，故\*5

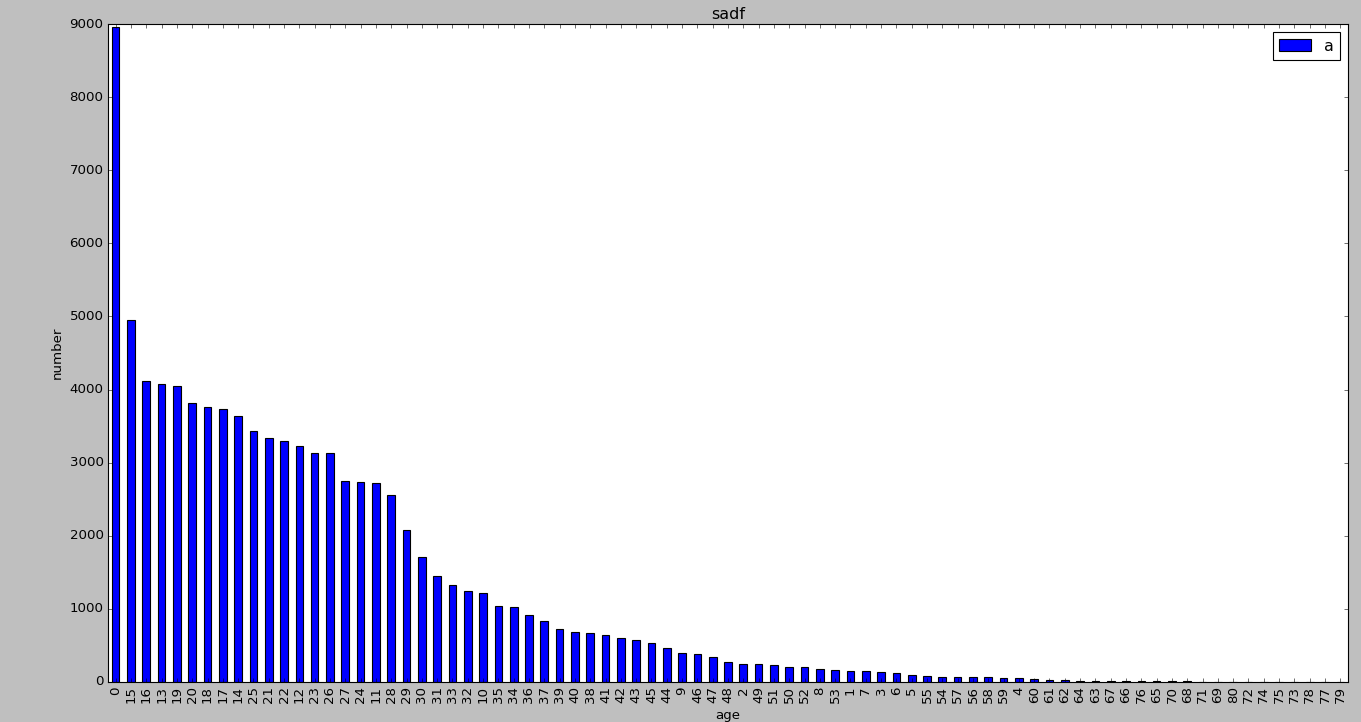
1、教育

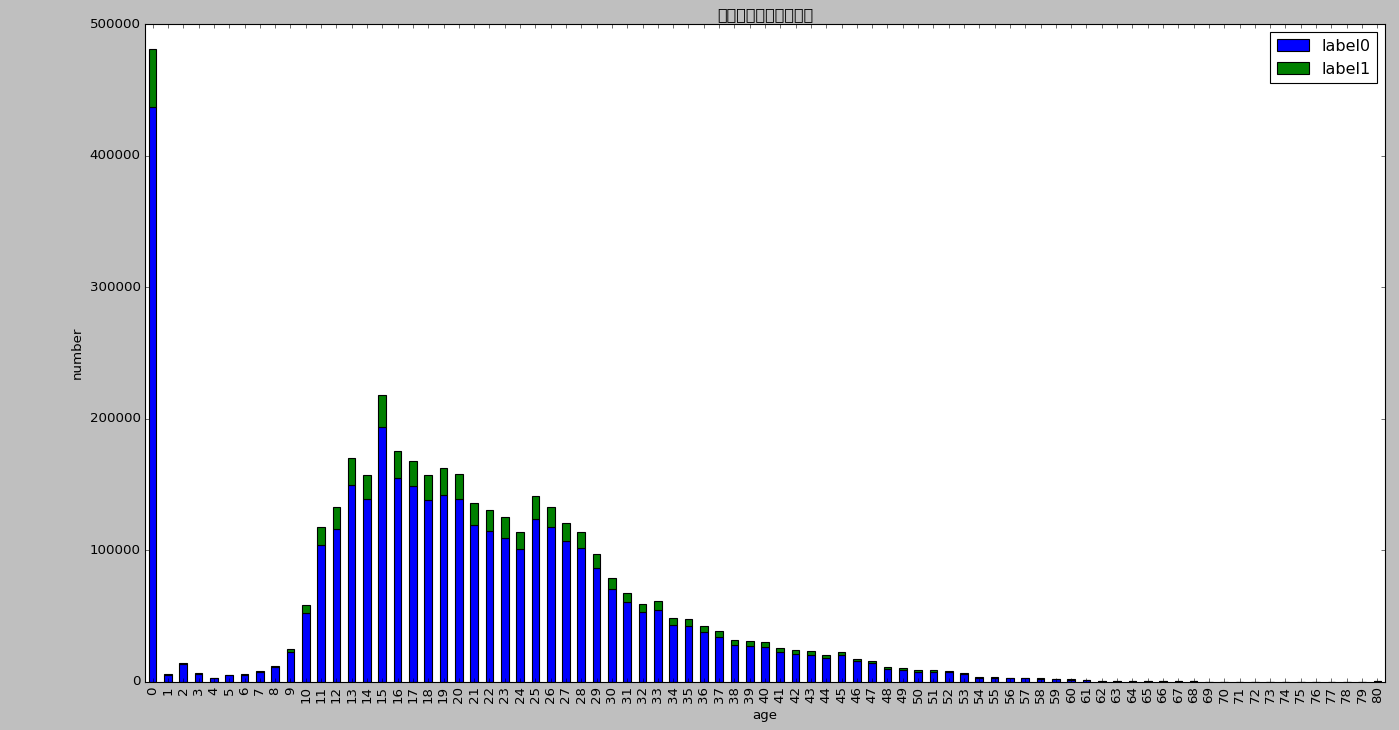




未知的太多（0），应考虑重新将未知的重新处理

2、年龄



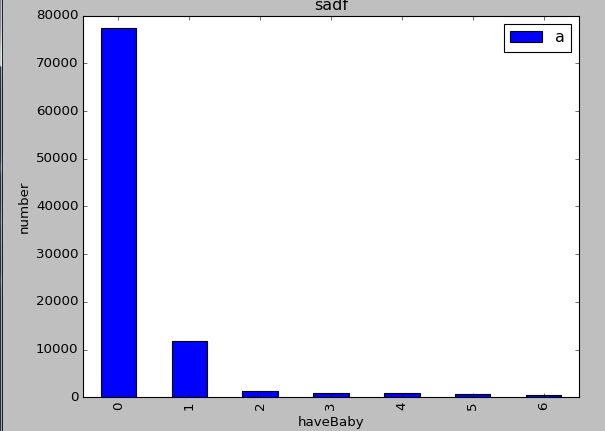


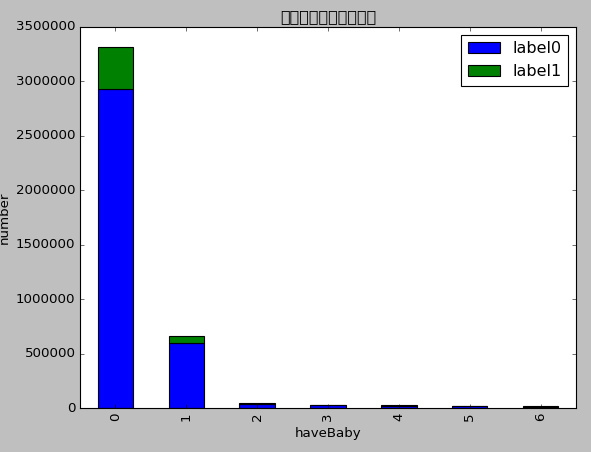
修改意见：1-8岁，50-80为一类，

9-10岁，30-49为一类，

11-29岁为一类

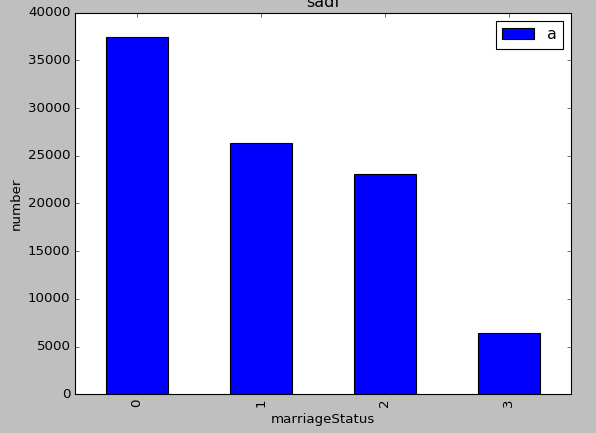
3、育儿

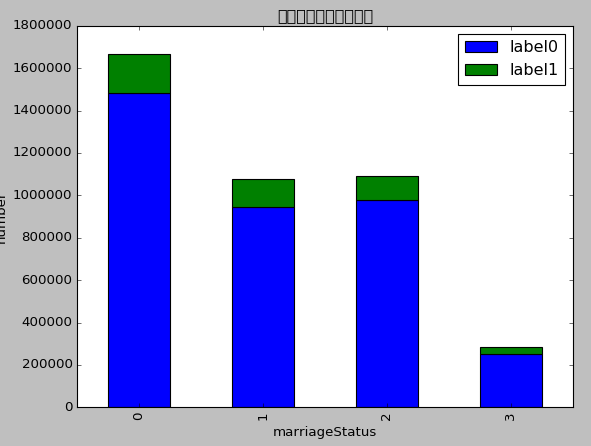




未知的太多，修改意见：要么对0重新赋值，要么考虑去除该特征

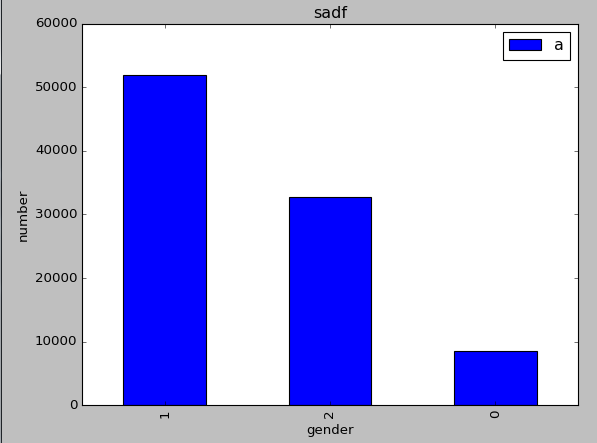
4、婚姻状况

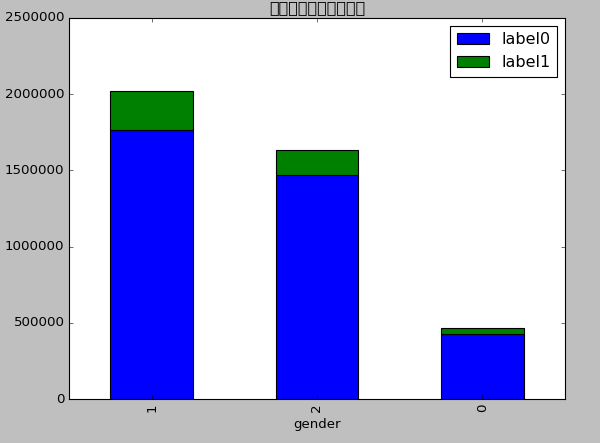




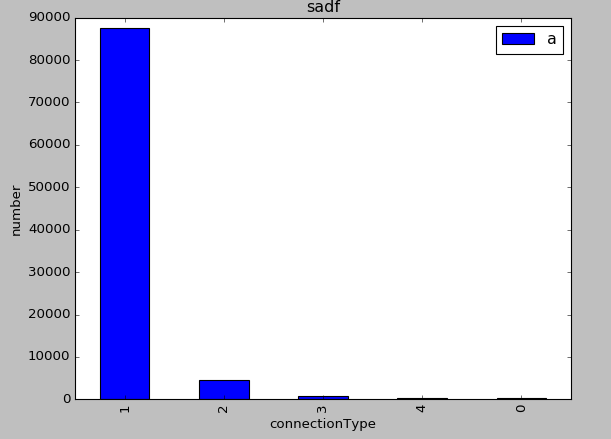
已经去除该特征，但是从图中看还可以接受，该如何处理

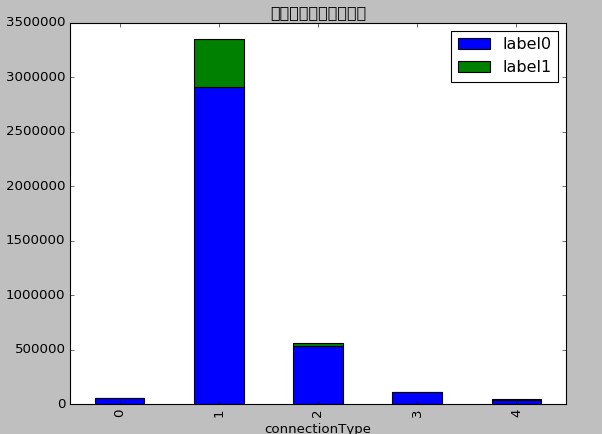
5、性别



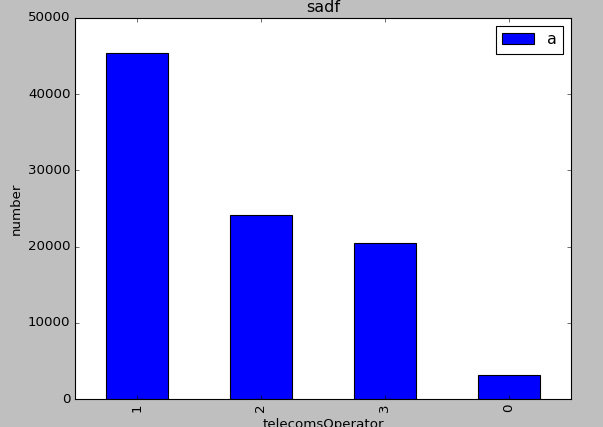


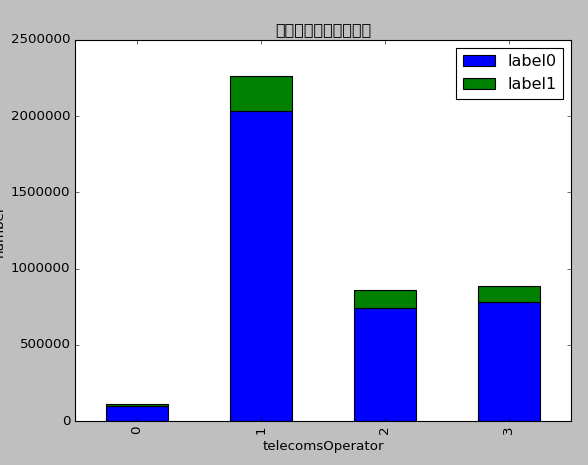
6、连接方式





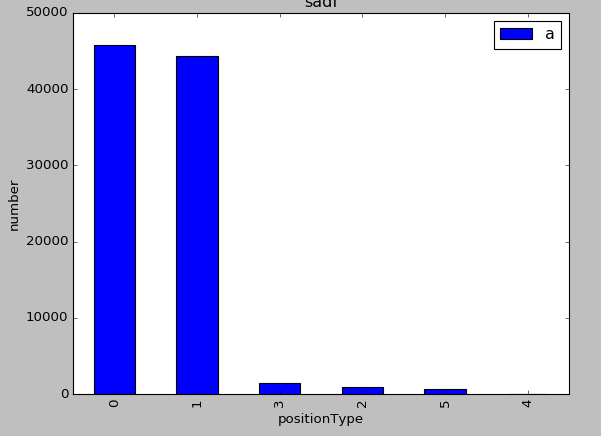
7、telecomsOperator

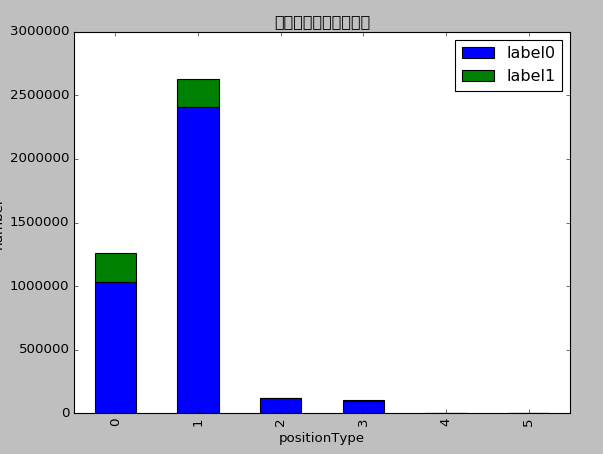




已经去除该特征，但是从图中看还可以接受，该如何处理

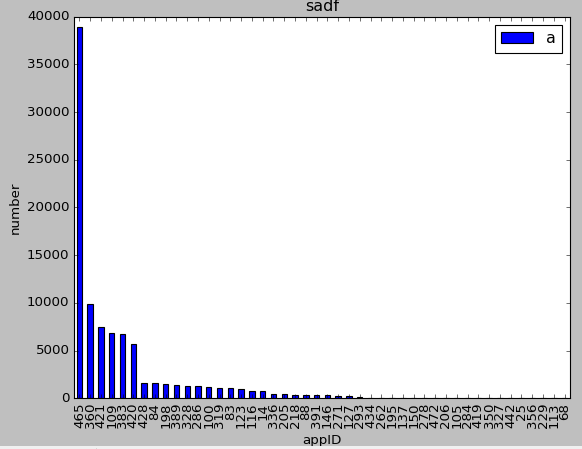
8、positionType

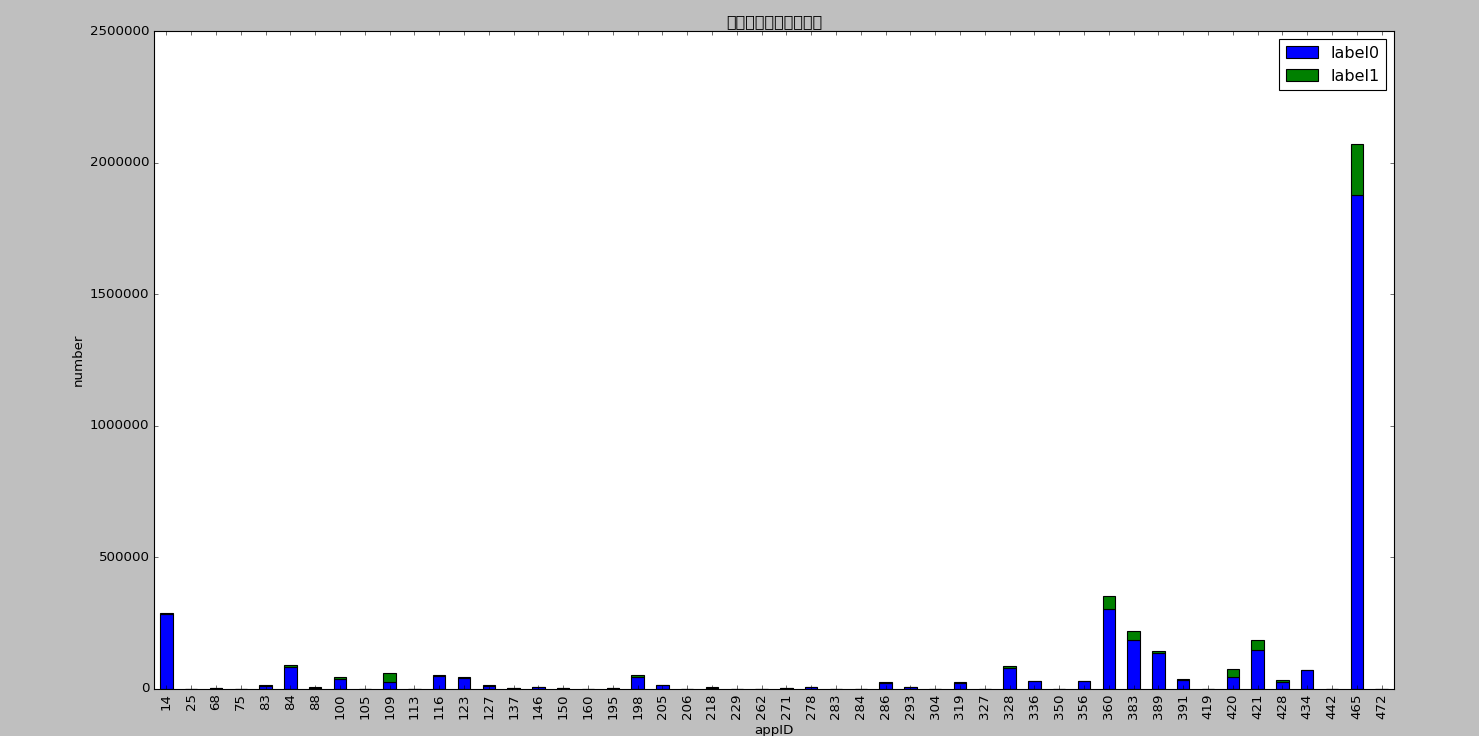




未知的太多（0），应考虑重新将未知的重新处理

9、appID





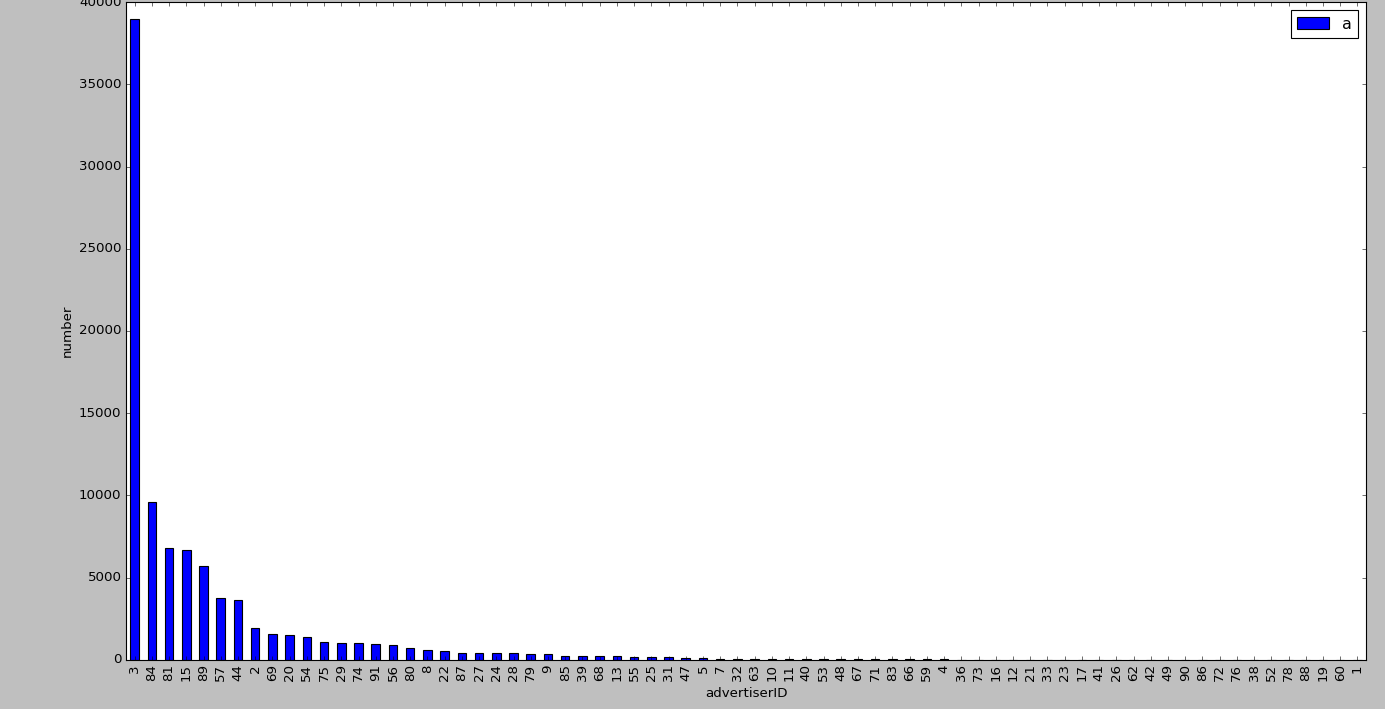
修改方案：109一类

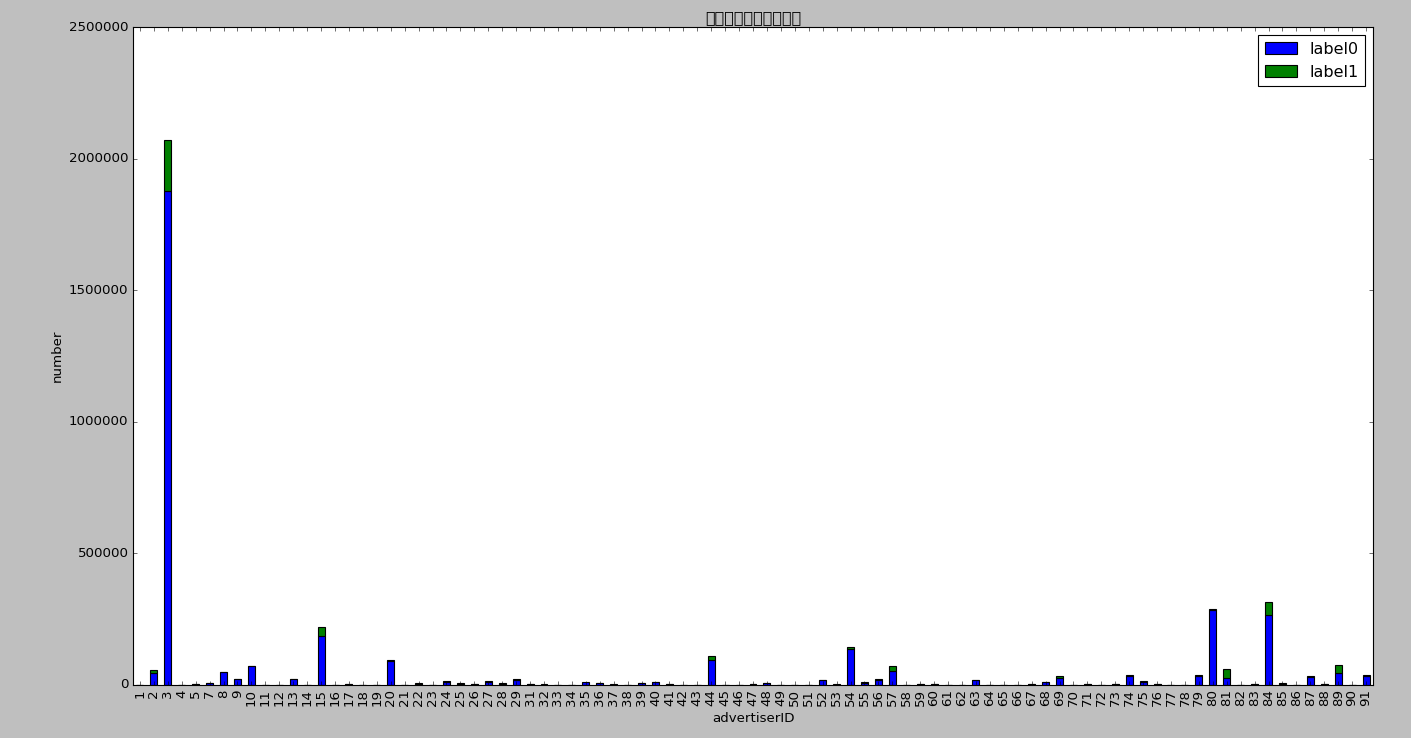
465，一类

360,421,,383,420一类

其他一类

10、advertiserID





修改方案：81,89一类

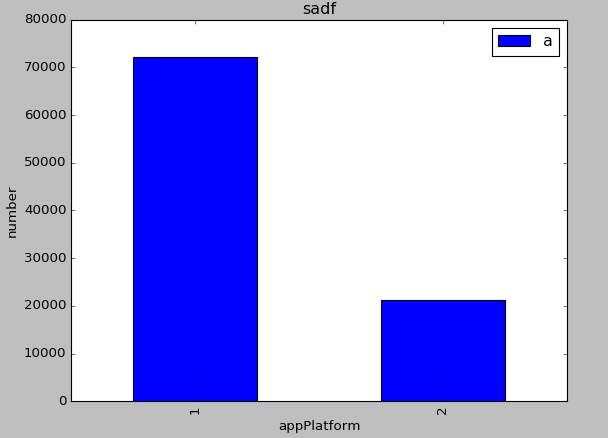
3一类

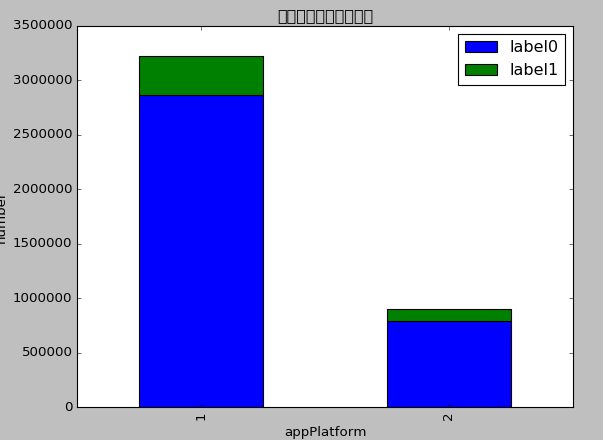
84, 15,57,4一类

2,69,20,54,75,29,74,91,56,80,8,22,87,27,24,28,79,9

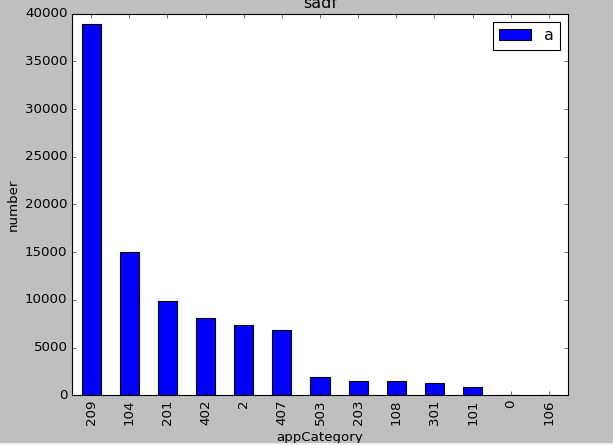
其他一类

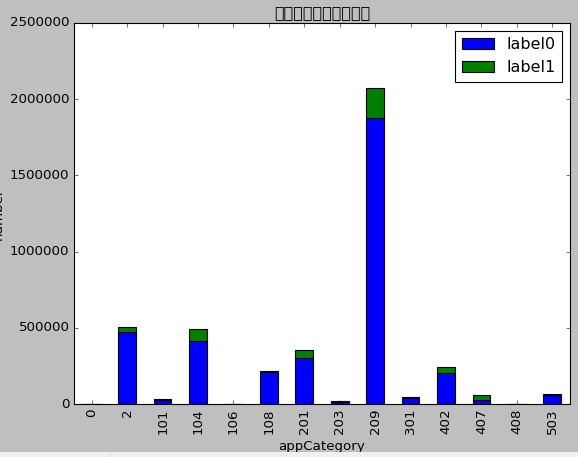
11、appPlatform





12、appCategory





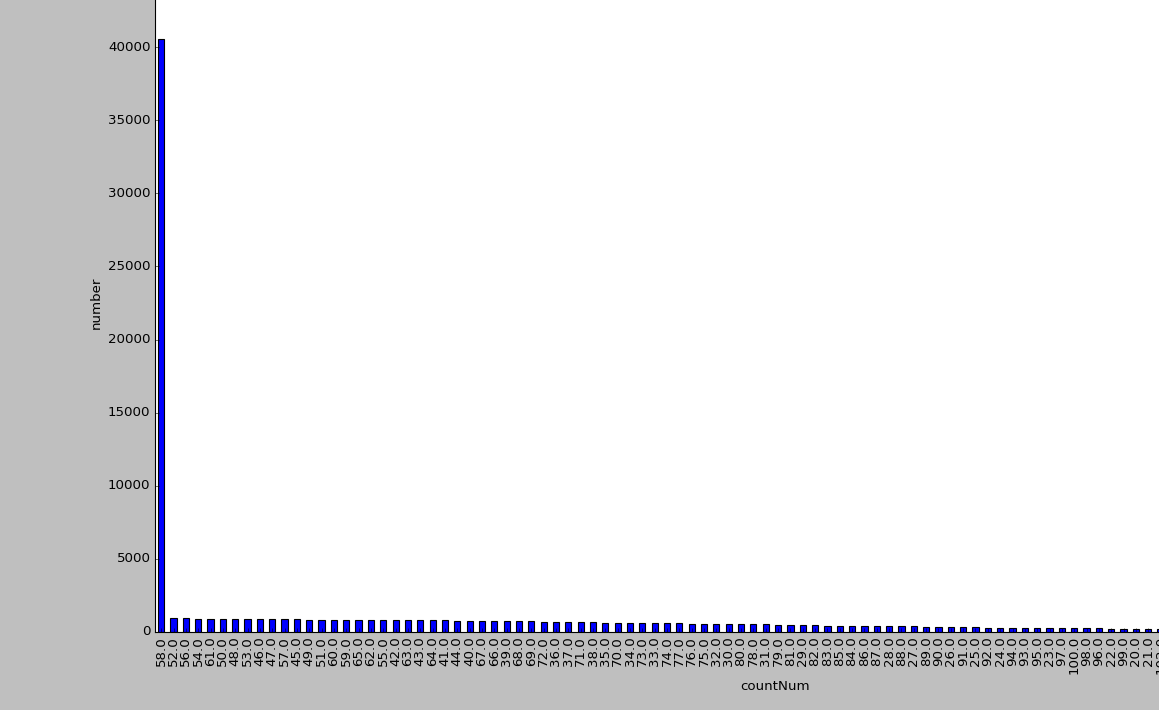
修改方案：209一类,407一类

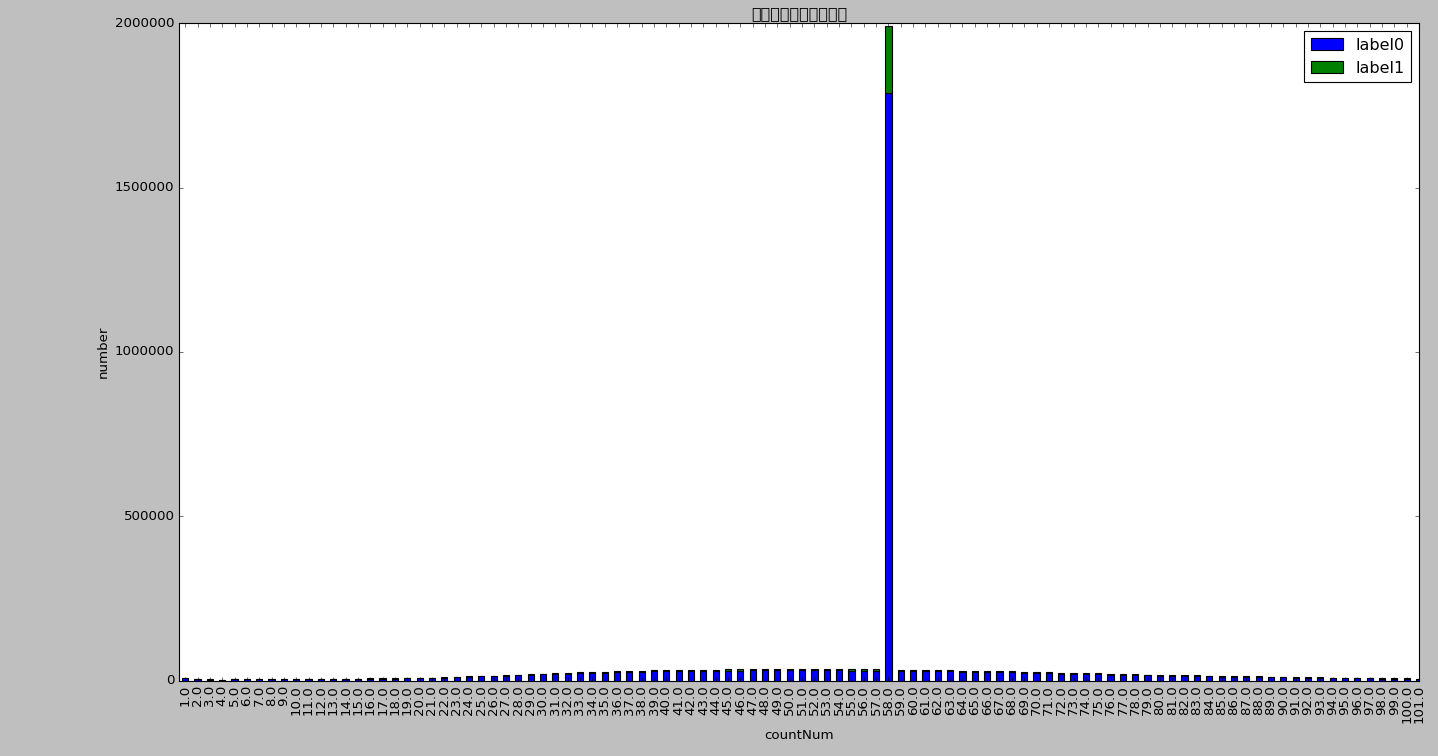
104,201,402,2一类

503,203,108,301,101一类

0,106一类

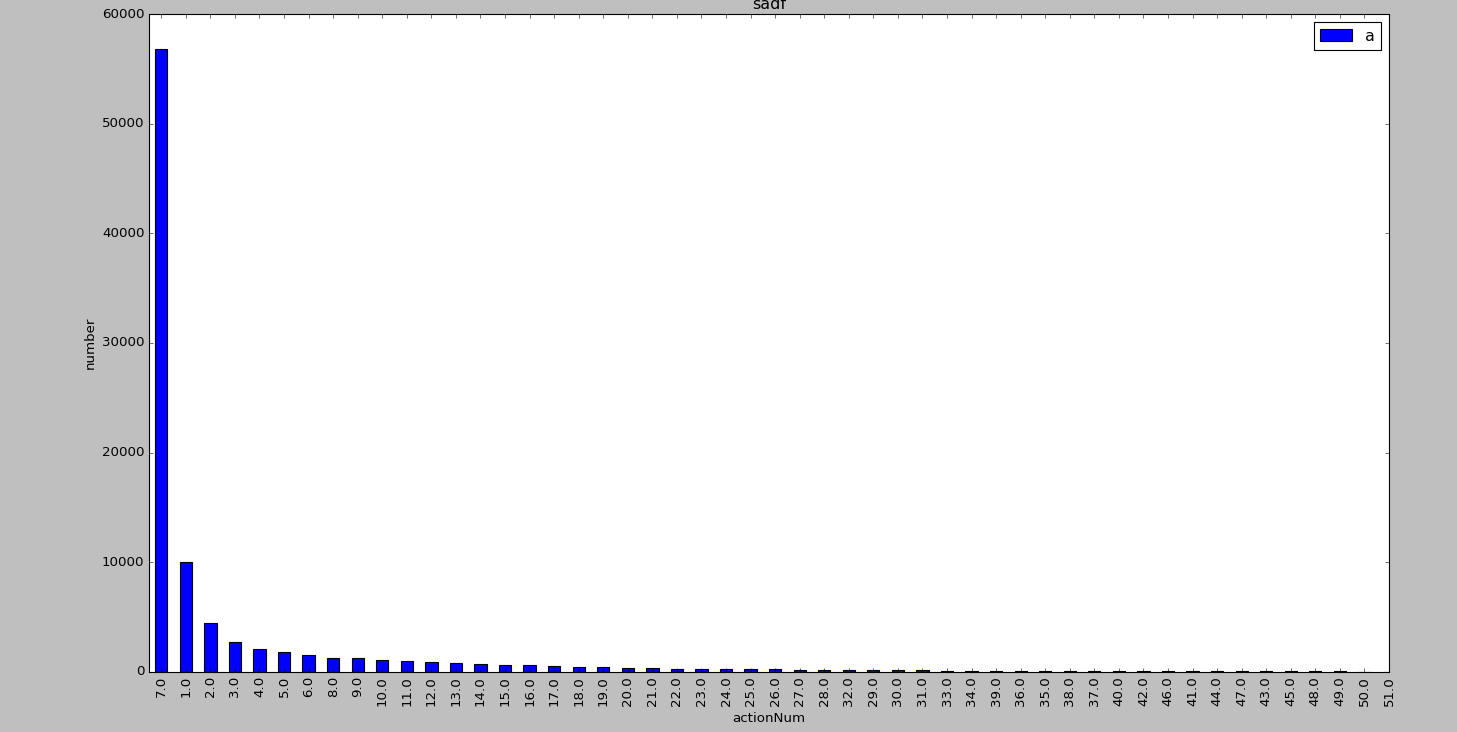
13、countNum

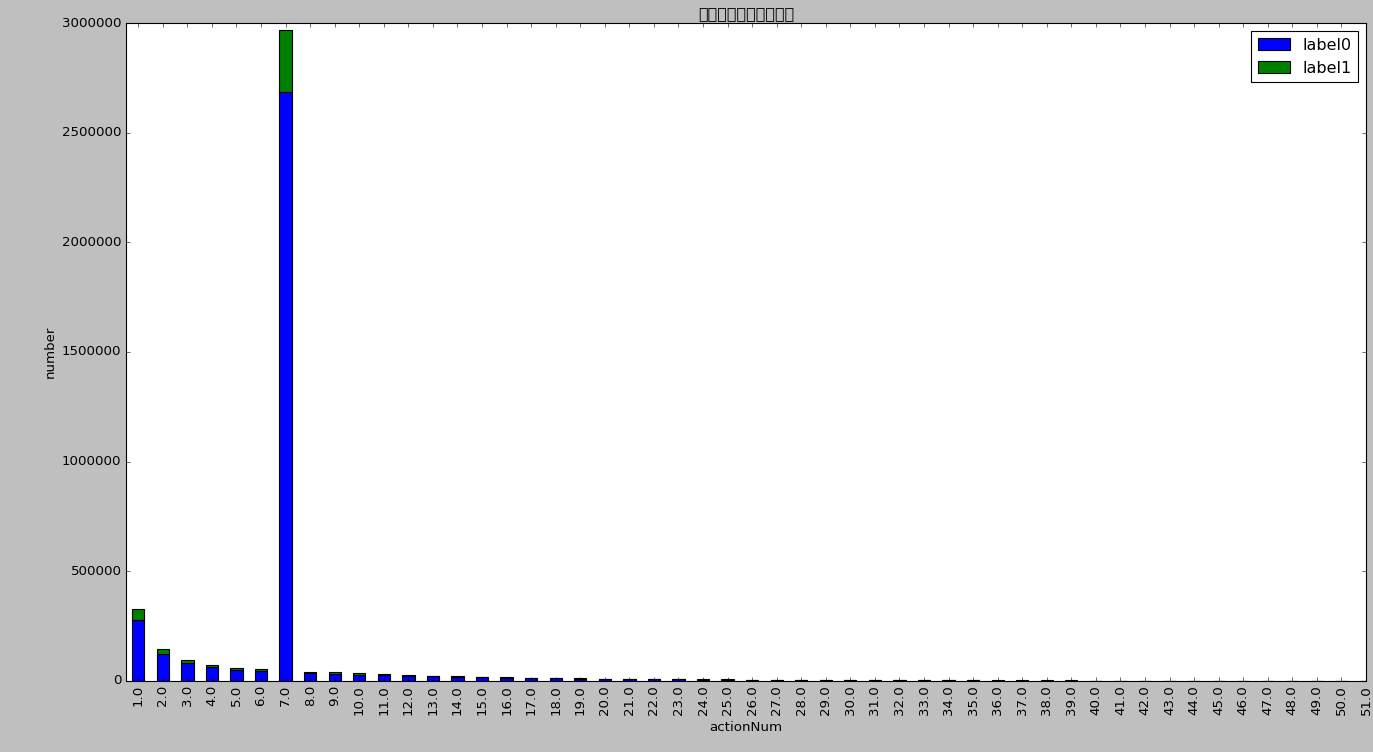




用随机森林重新给空缺值进行赋值

14、actionNum





用随机森林重新给空缺值进行赋值

