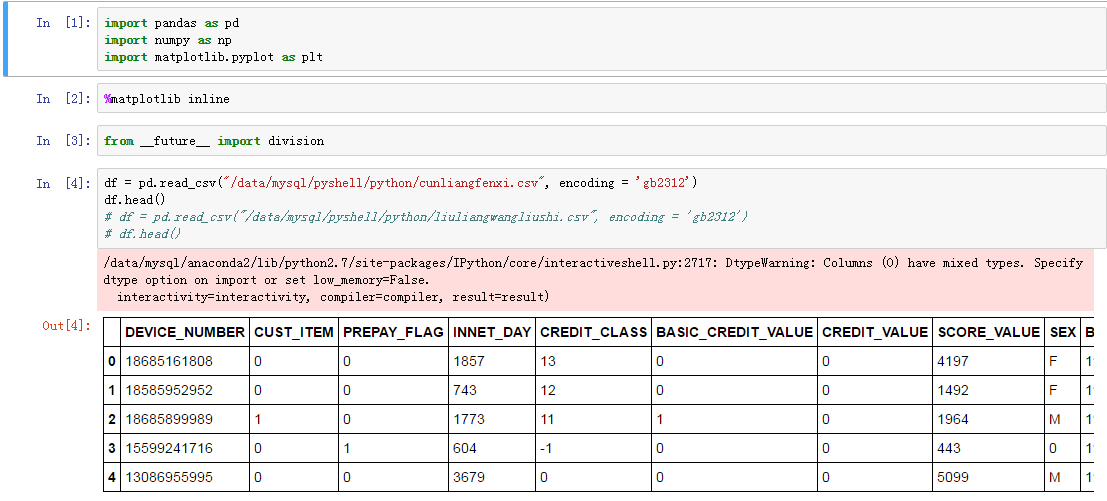
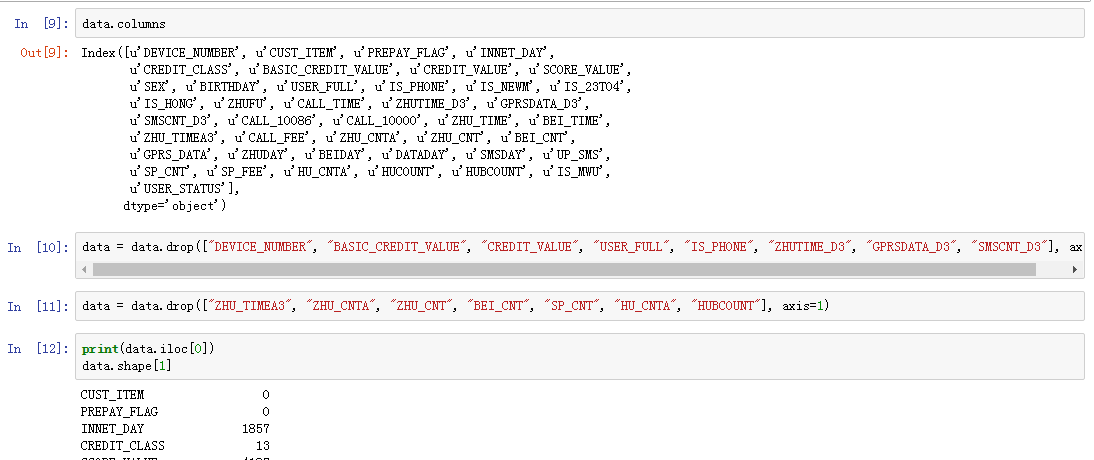
1. 存量分析



读取CSV文件



去掉空缺大于一半以上的值



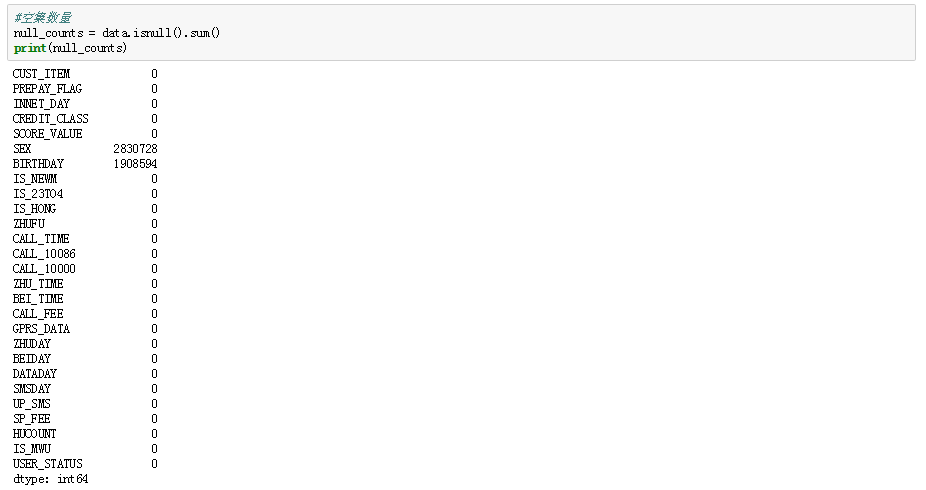
剔除与建模X值无关的值



显示y值



处理y值



显示空集数量



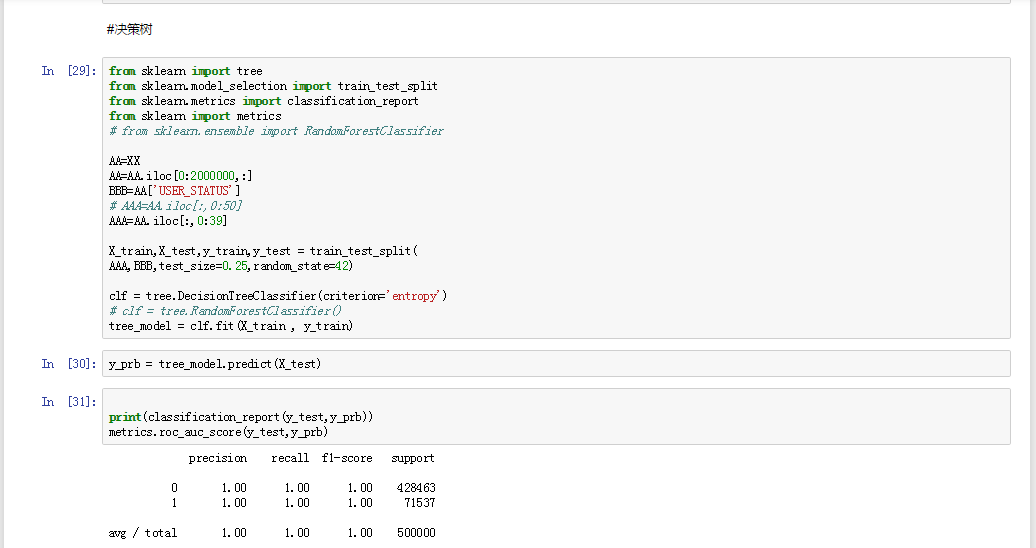
去掉空集并显示字段数据类型



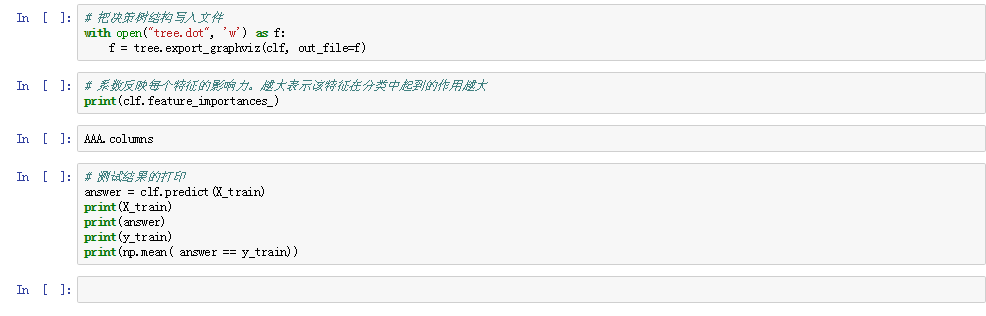
哑变量处理

建模：

1、决策树

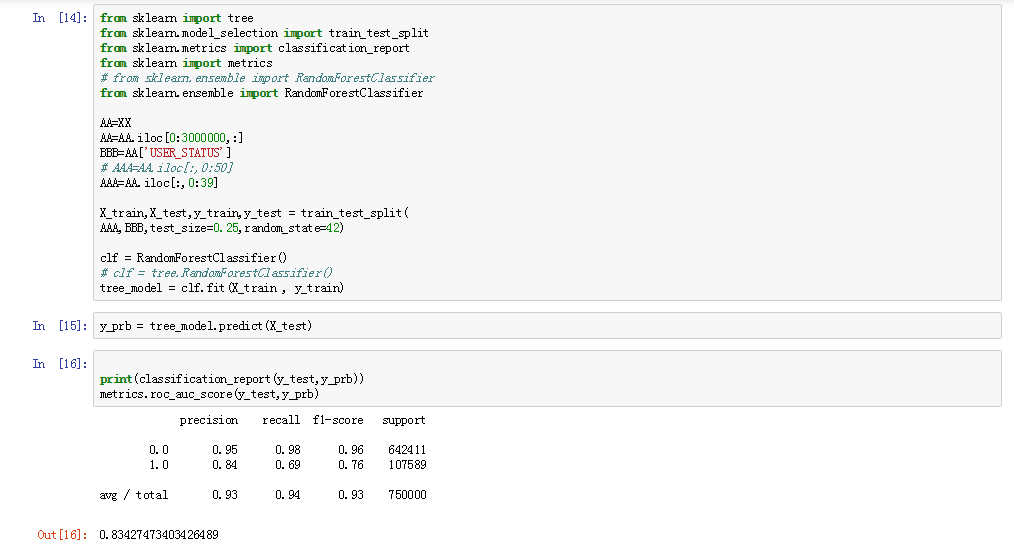


决策树建模



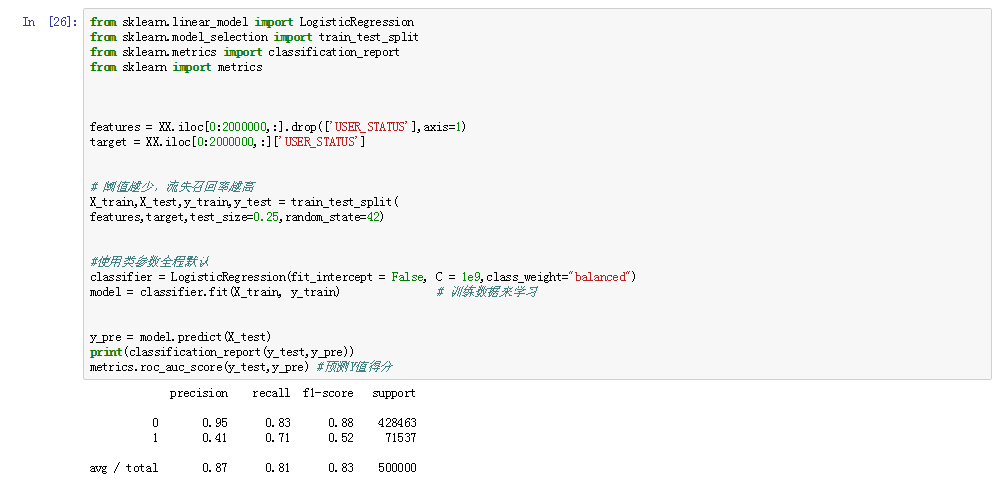
生产dot文件及信息增益

2、随机森林



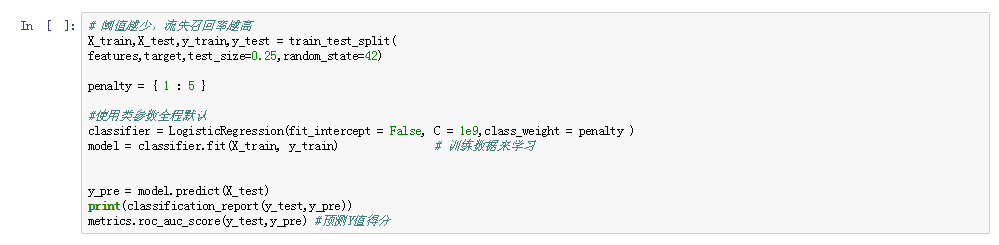
3、逻辑回归





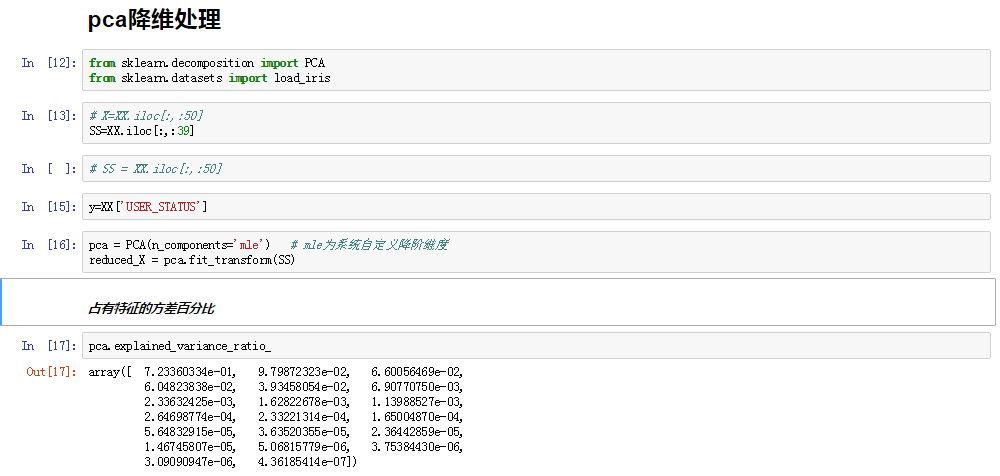


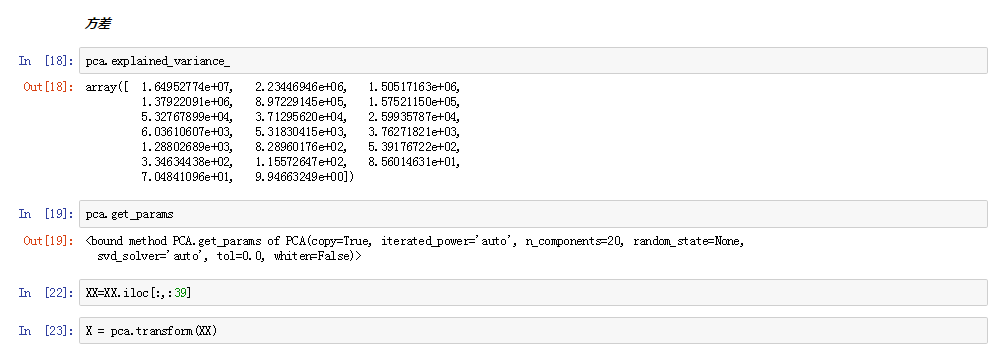
逻辑回归建模及生成增益



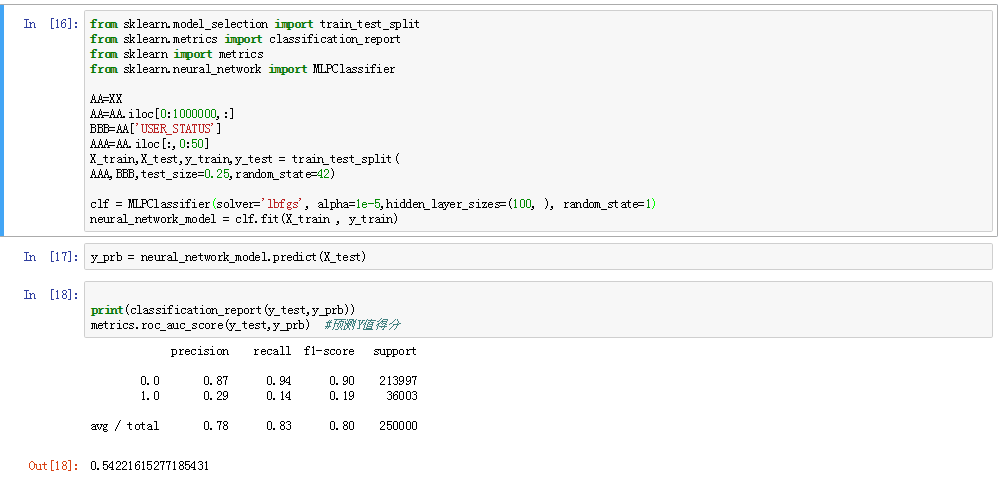
调整参数

4、PAC降维处理

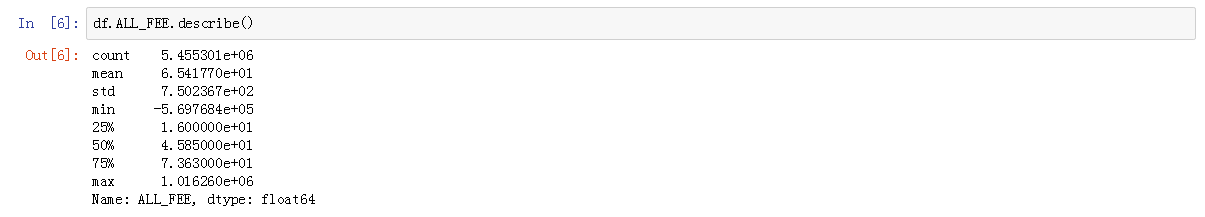




5、神经网络



6、描述性统计



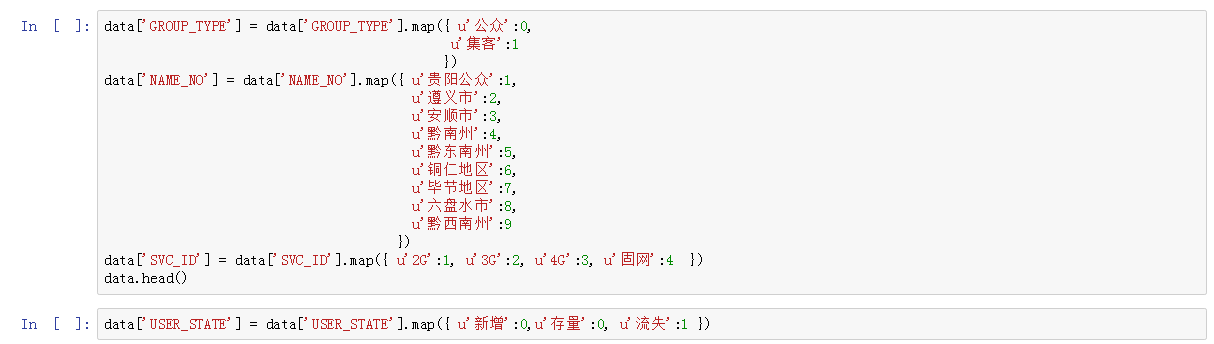
确定某一个字段，做描述性统计

其他数据清理方法：

列表的切割方法：



调用map字典做文字处理：



文字较多情况写geneum函数处理

