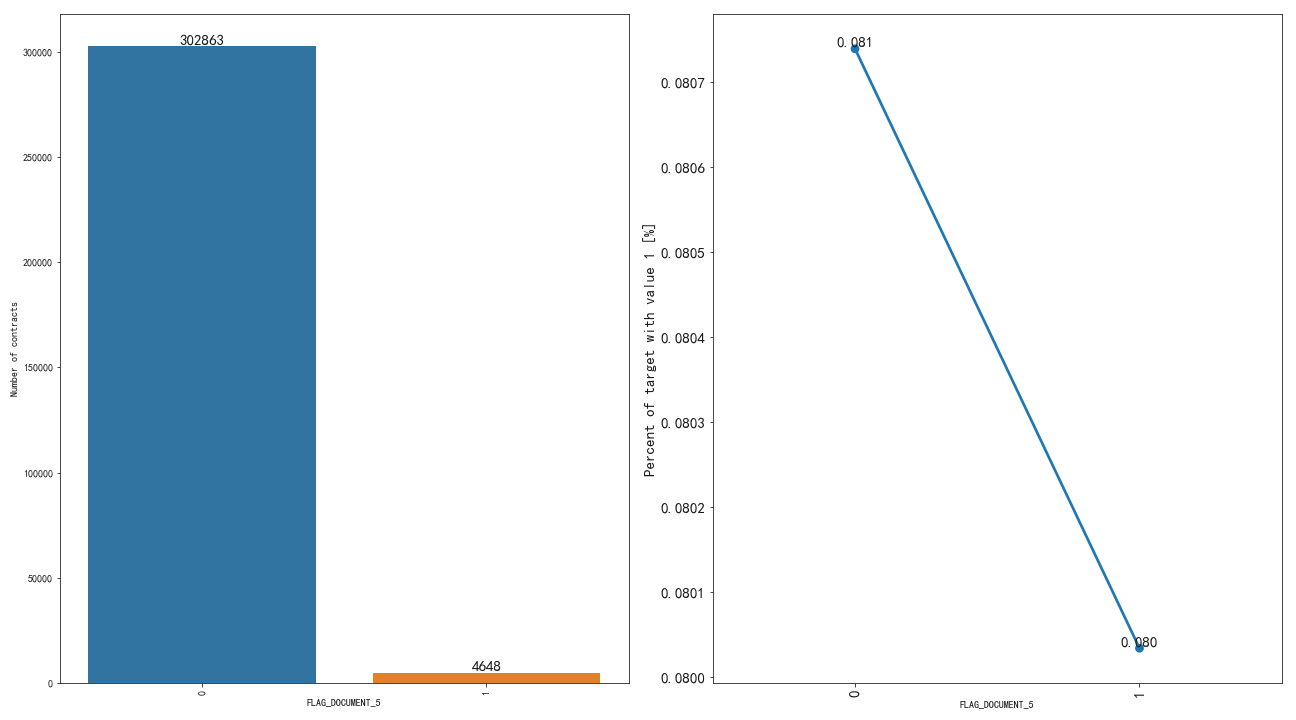
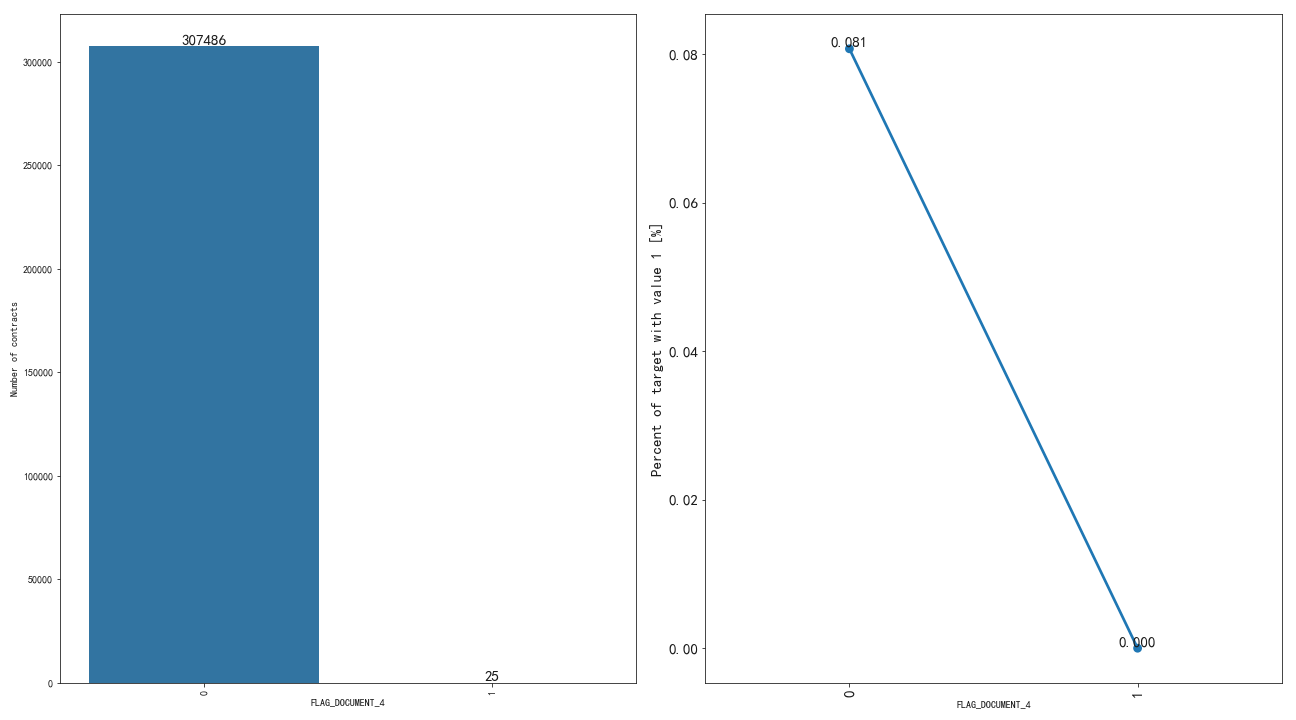
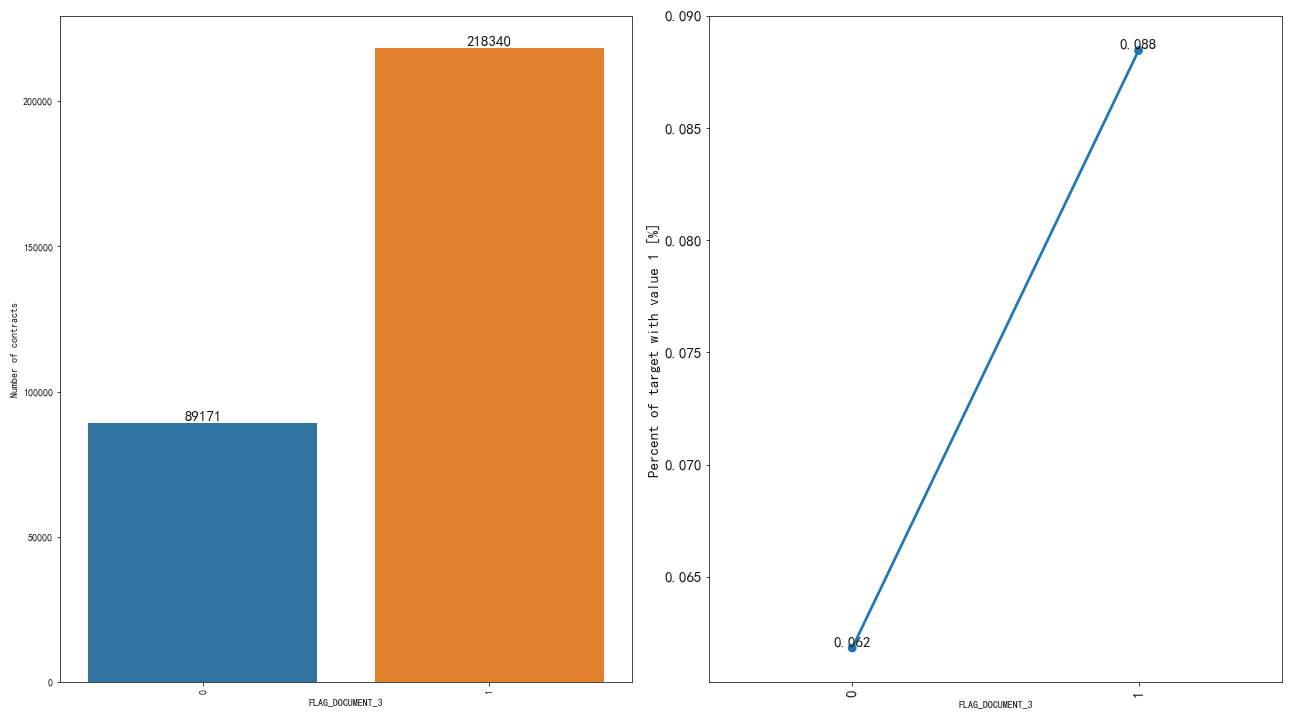
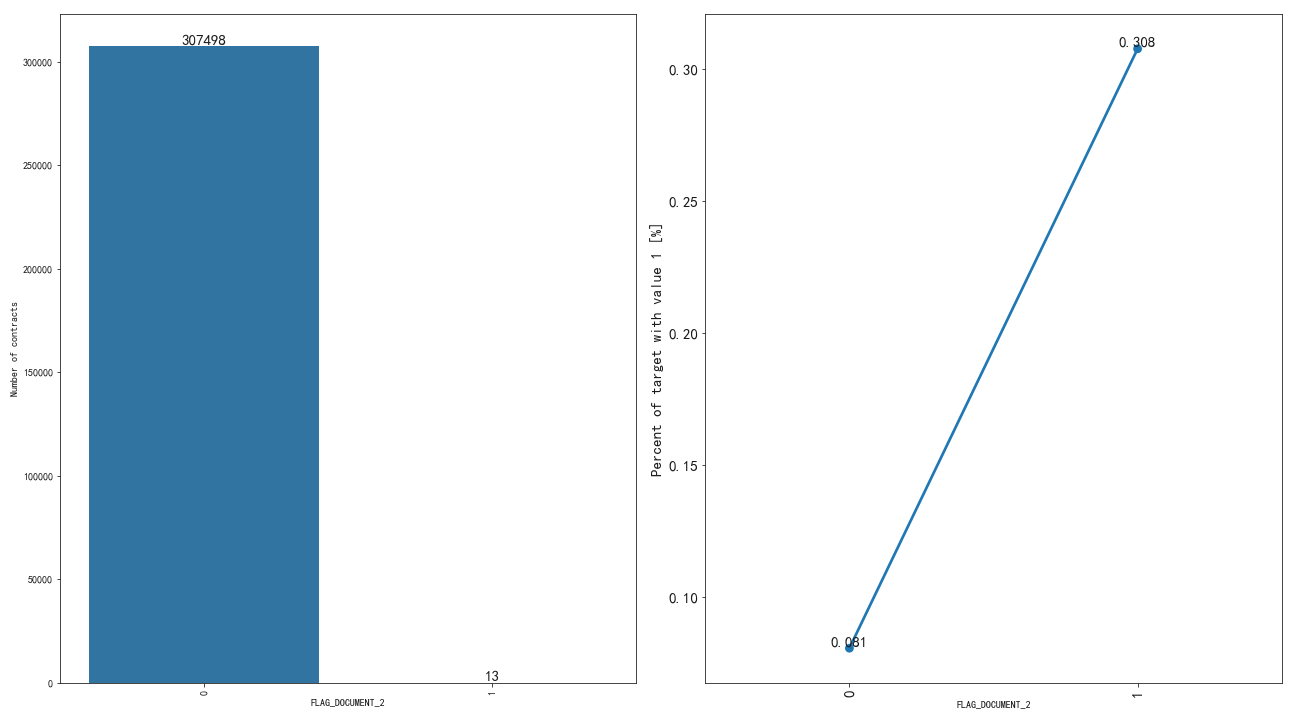
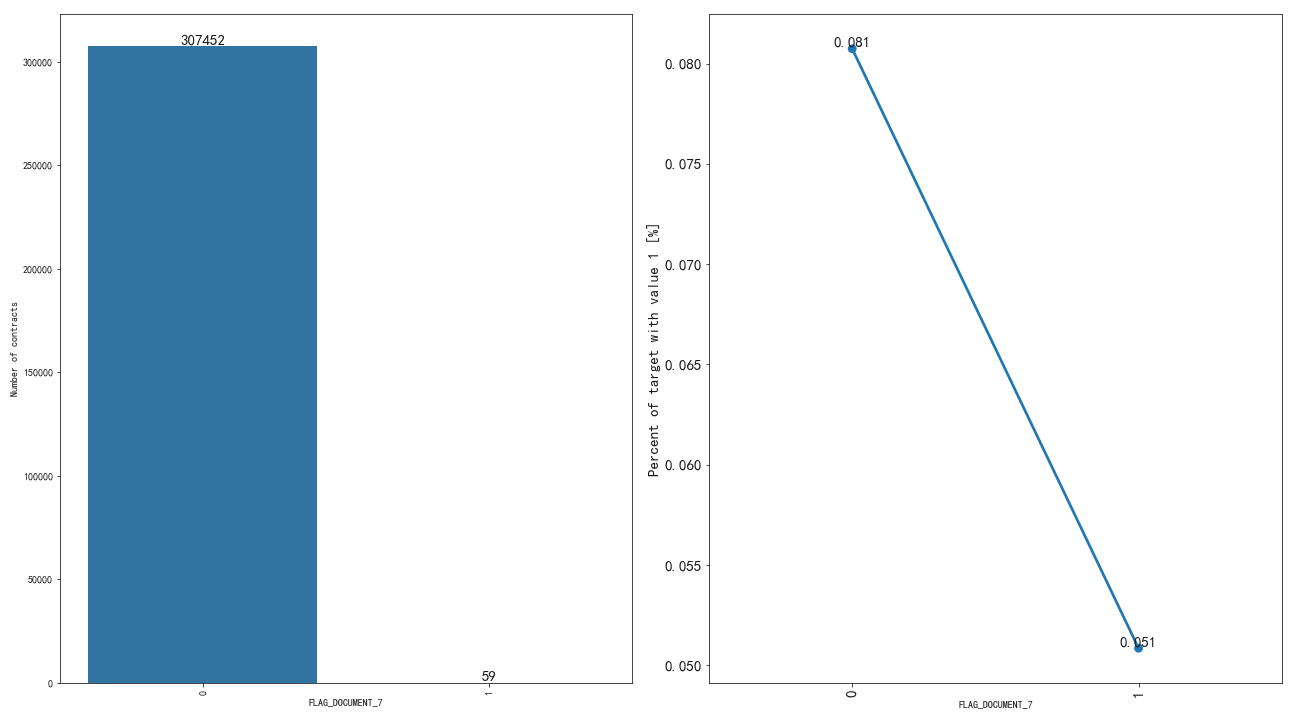
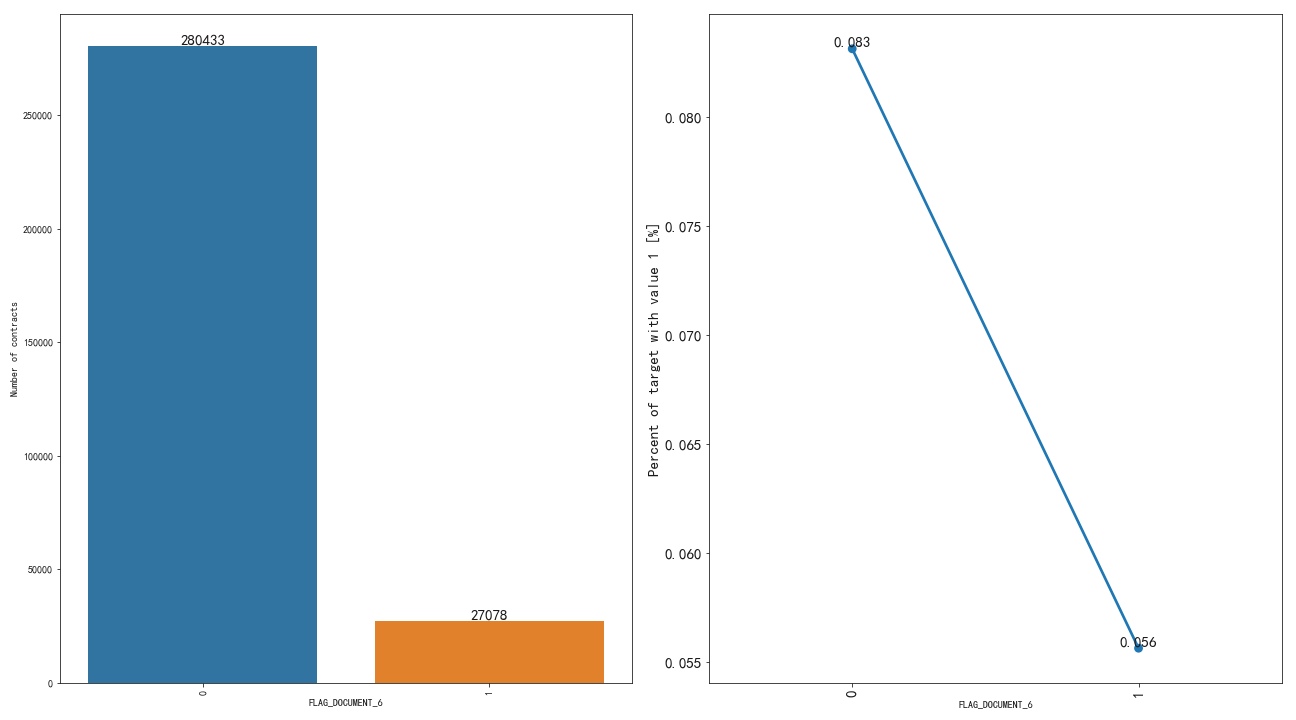
Main table 平均违约率为0.0807

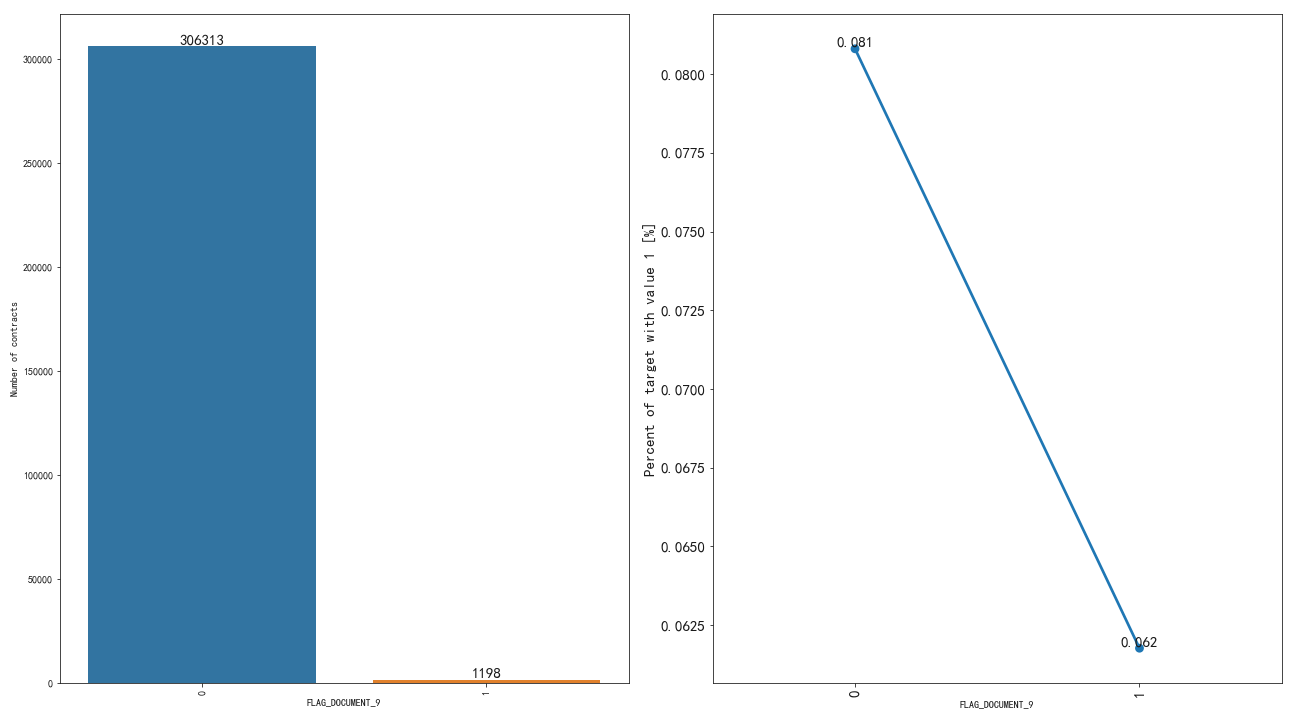
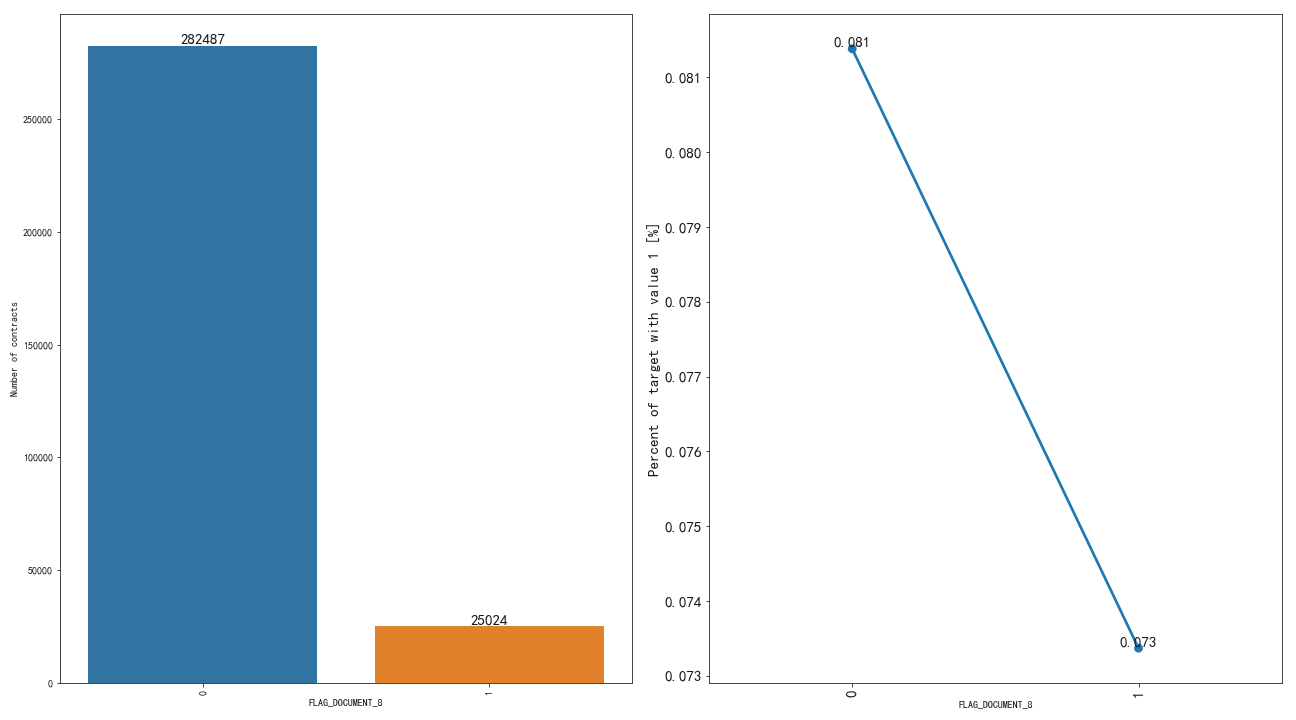
# 贷款人信息完整程度

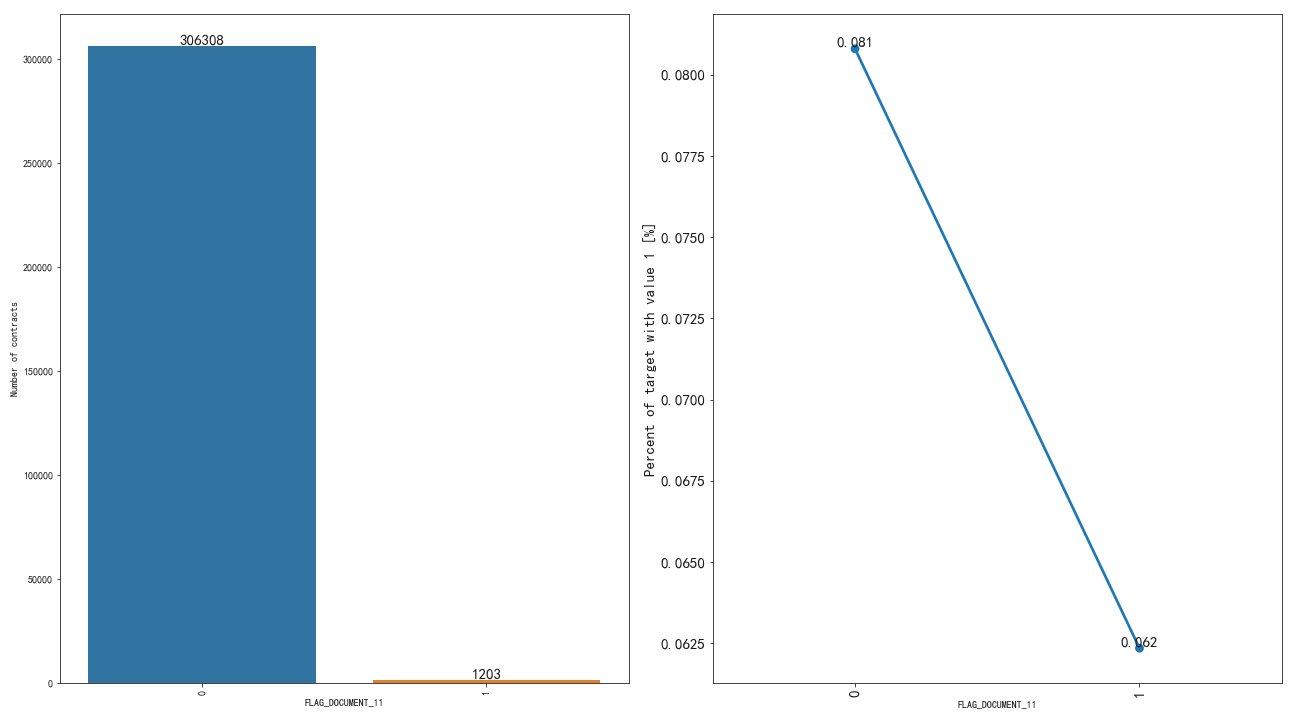
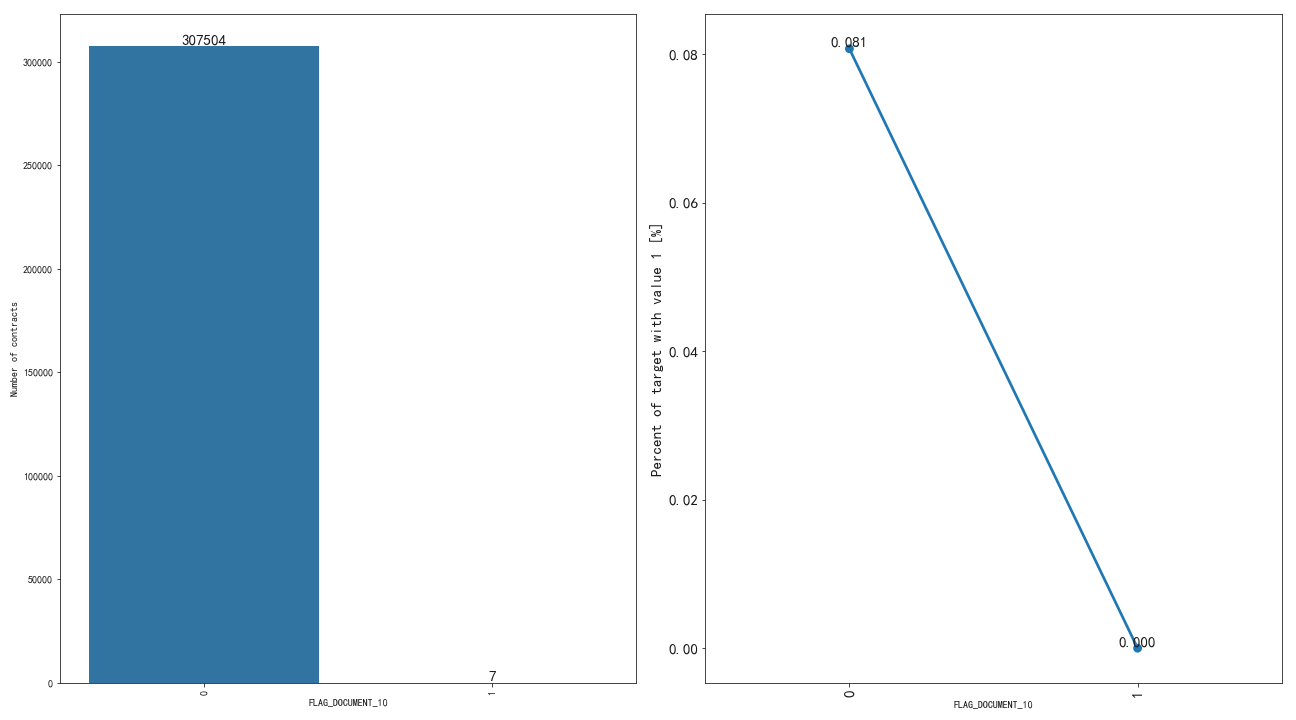
## 提交文件完整程度

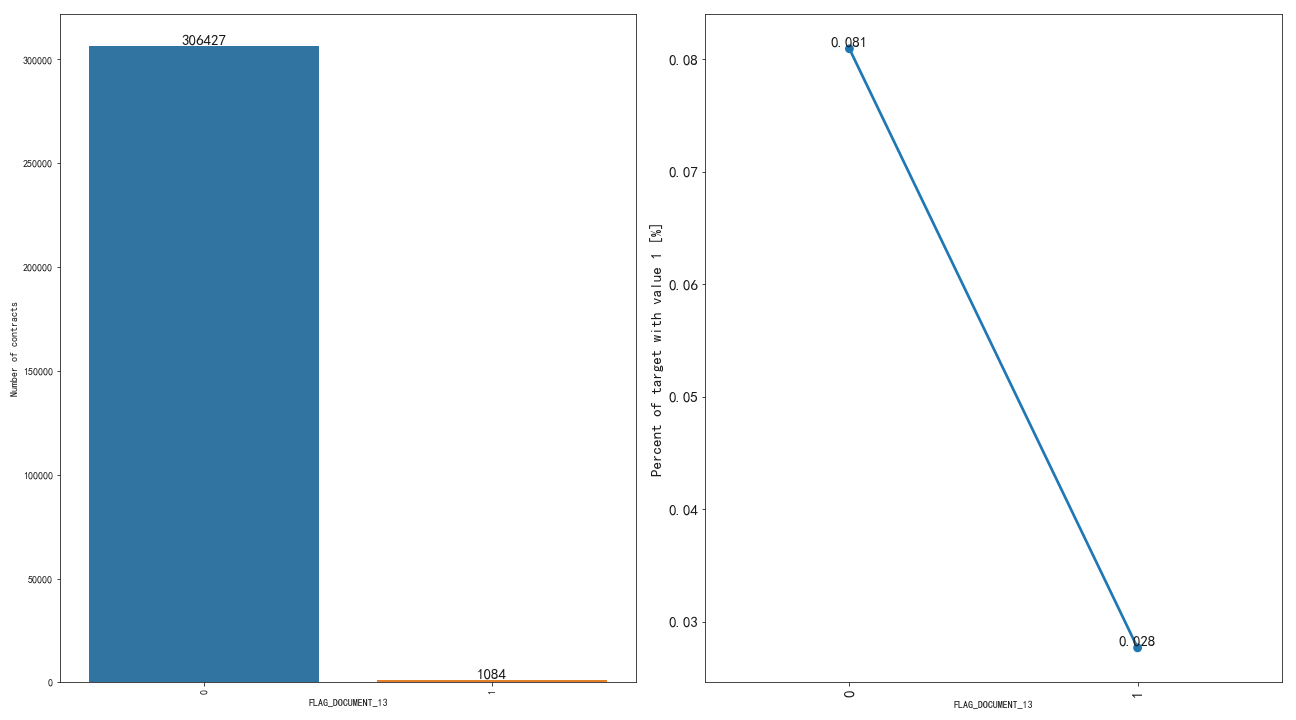
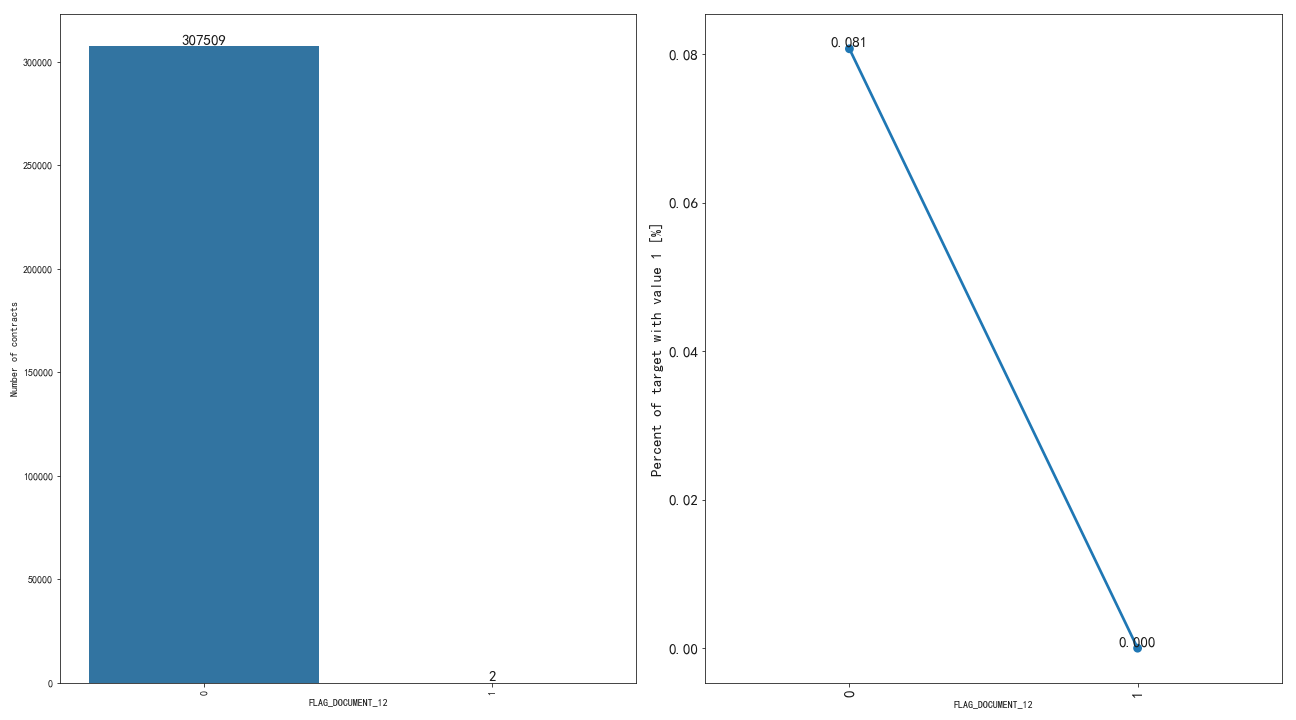
### FLAG\_DOCUMENT\_X(x=(2-21)

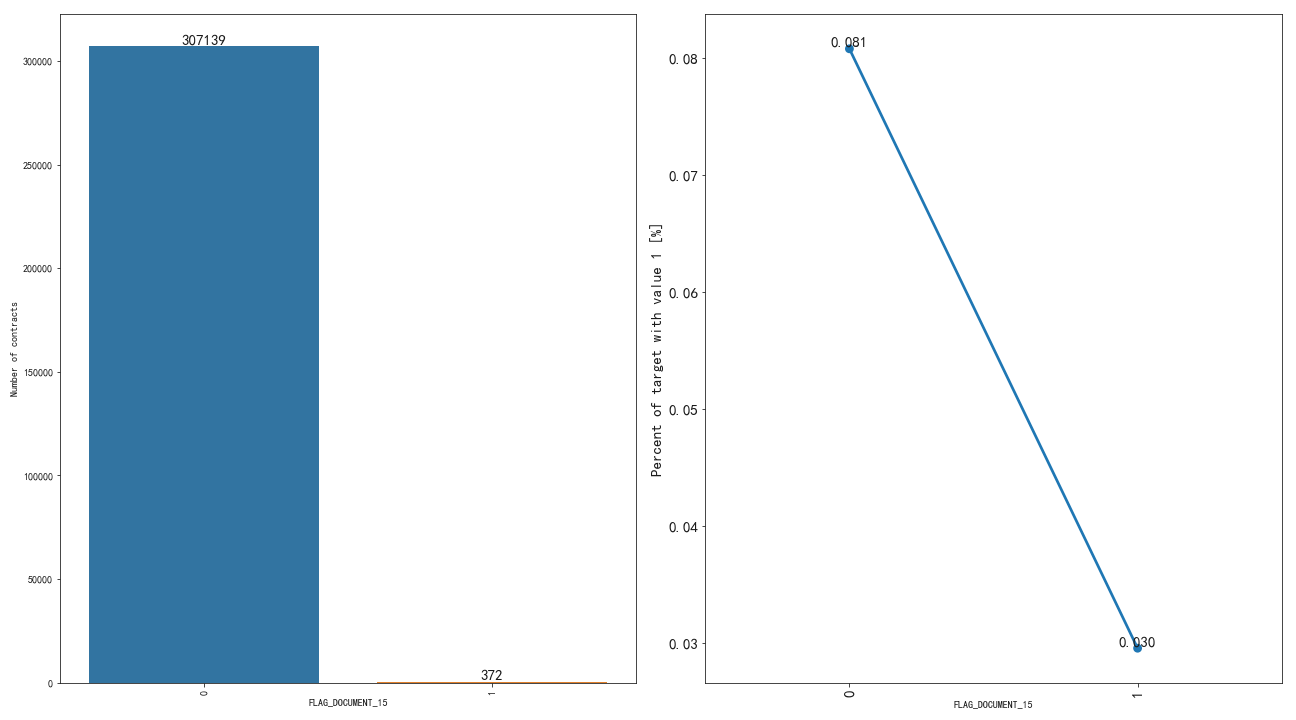
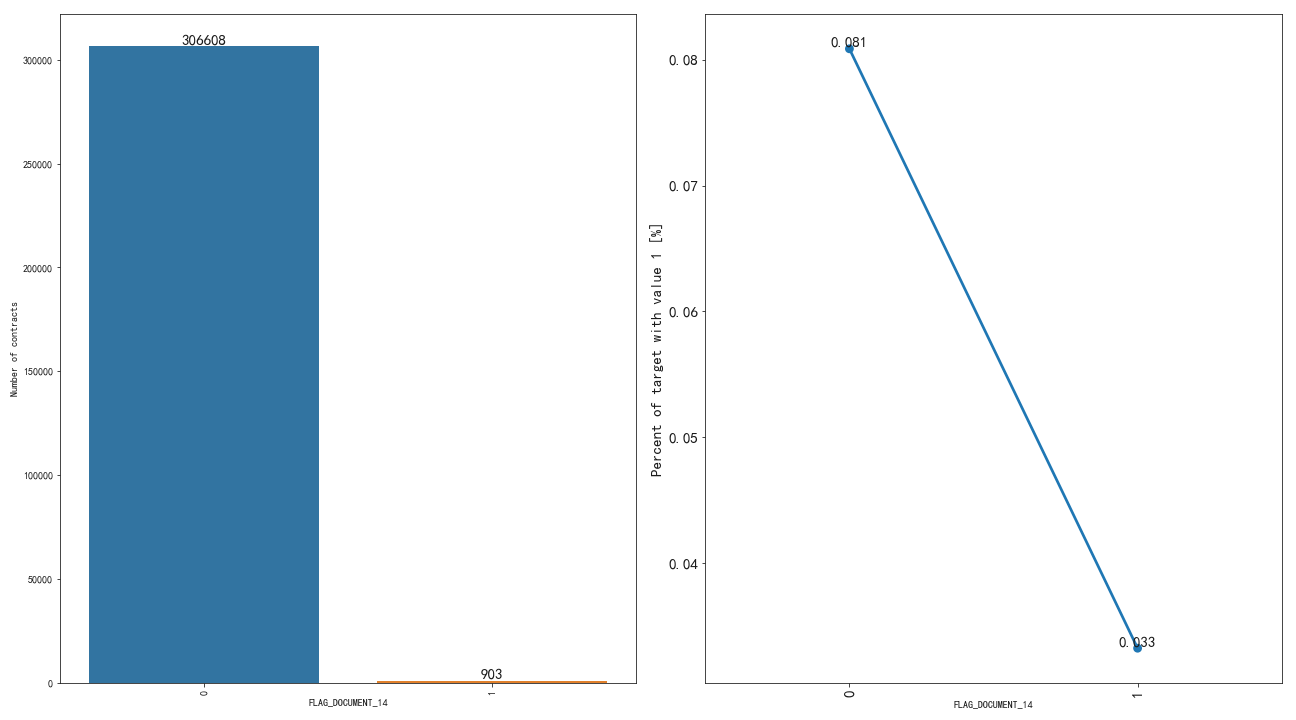


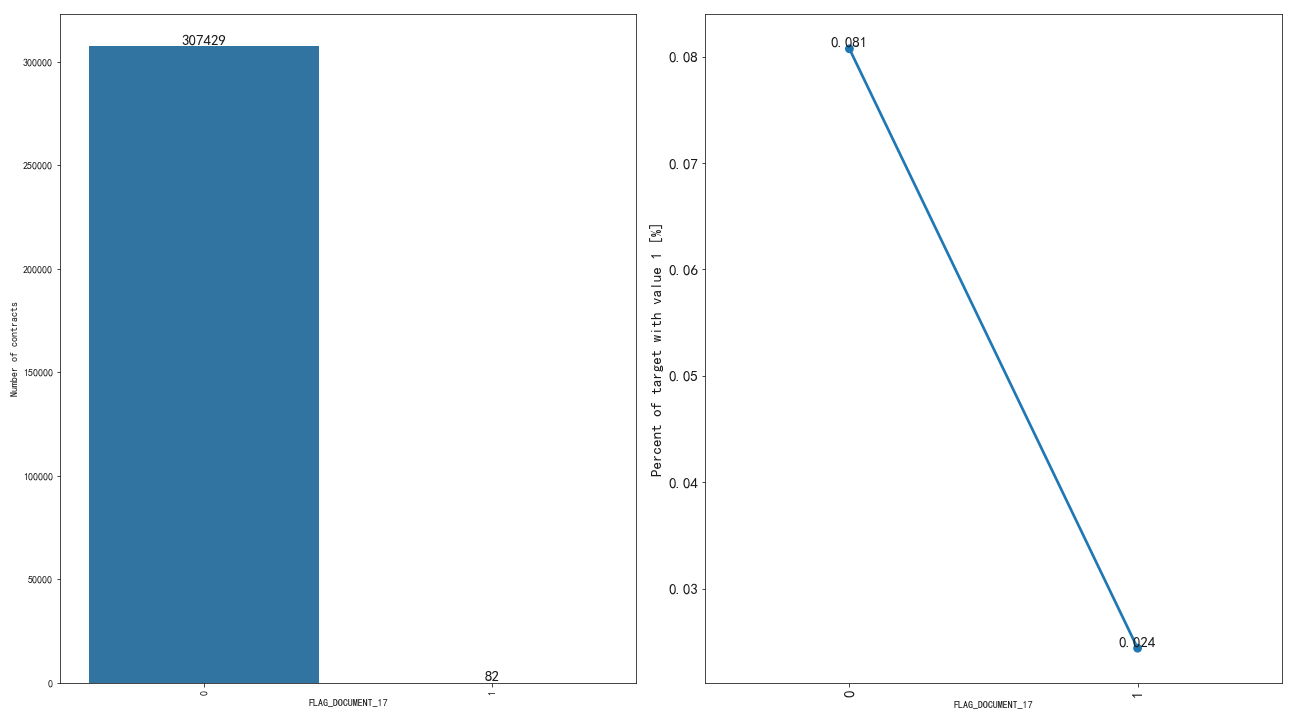
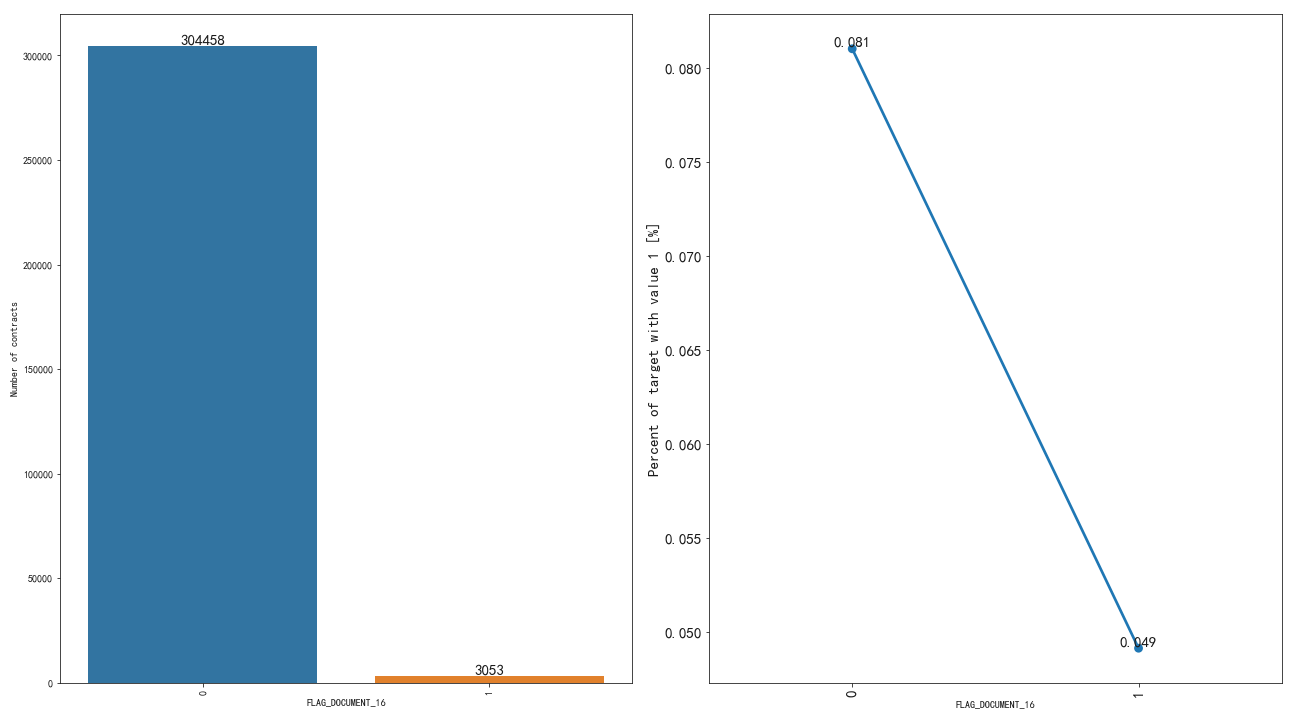


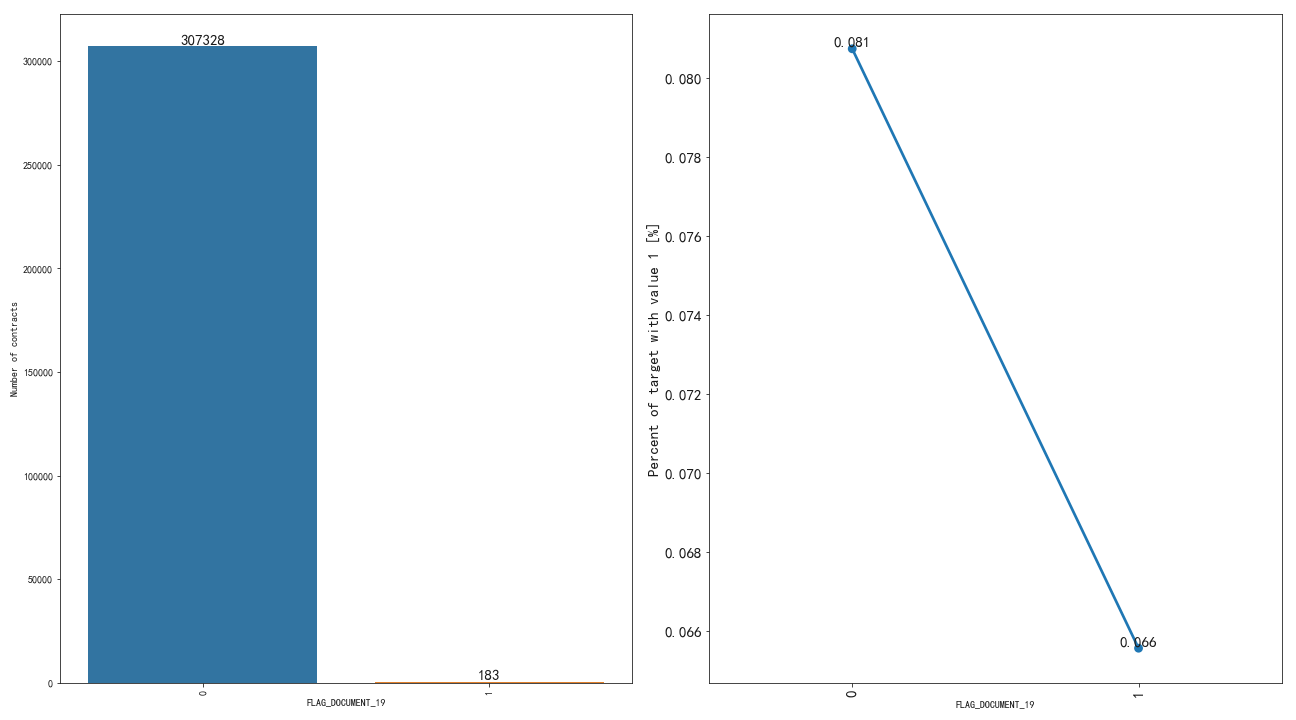
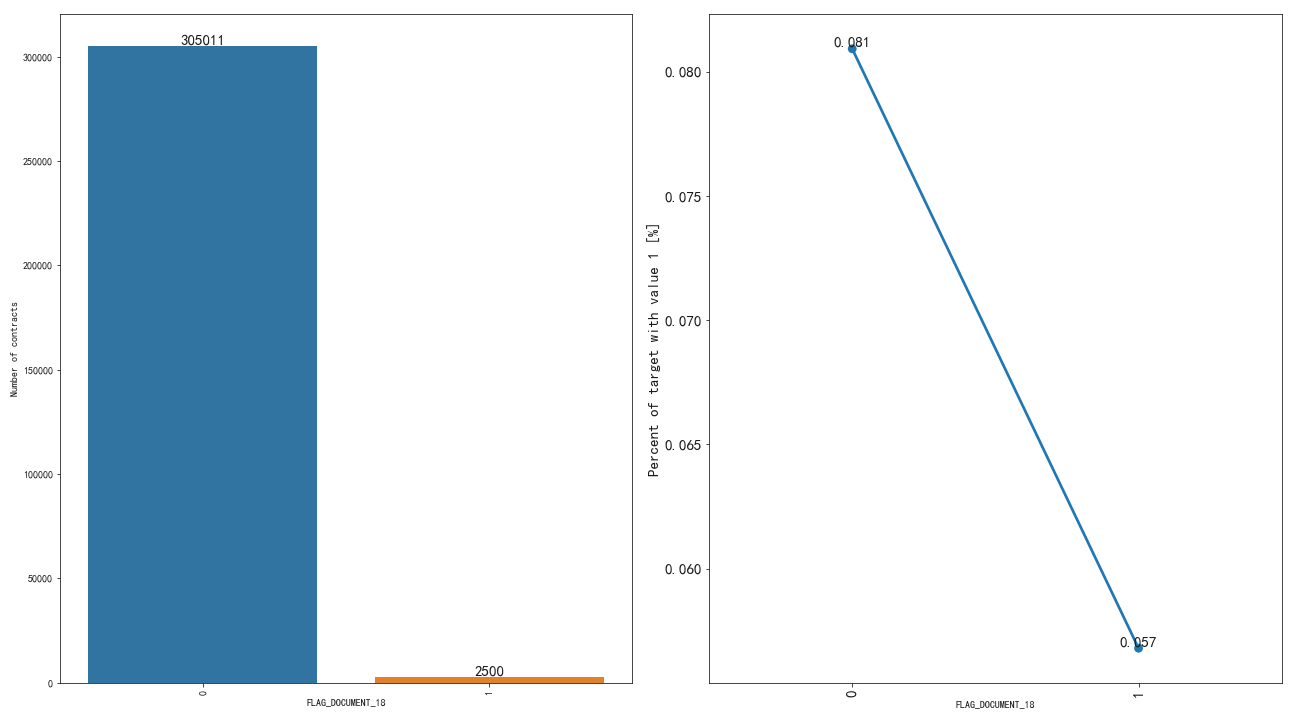


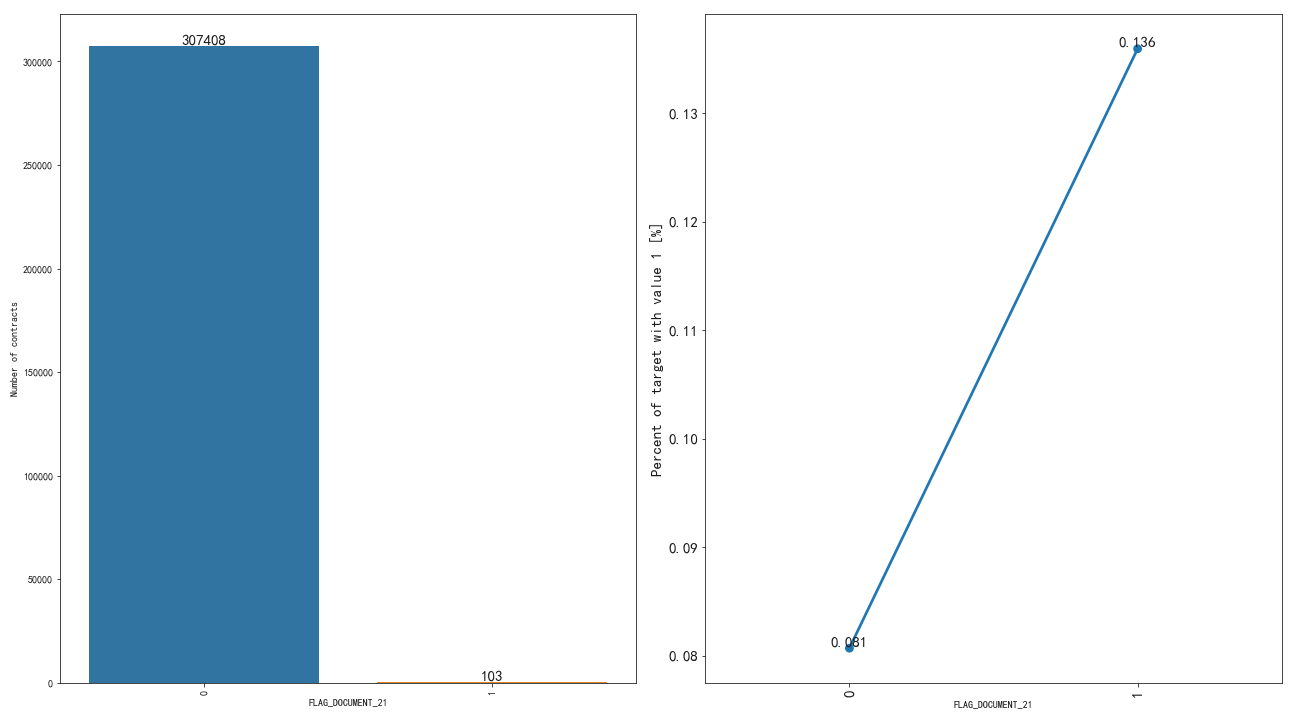
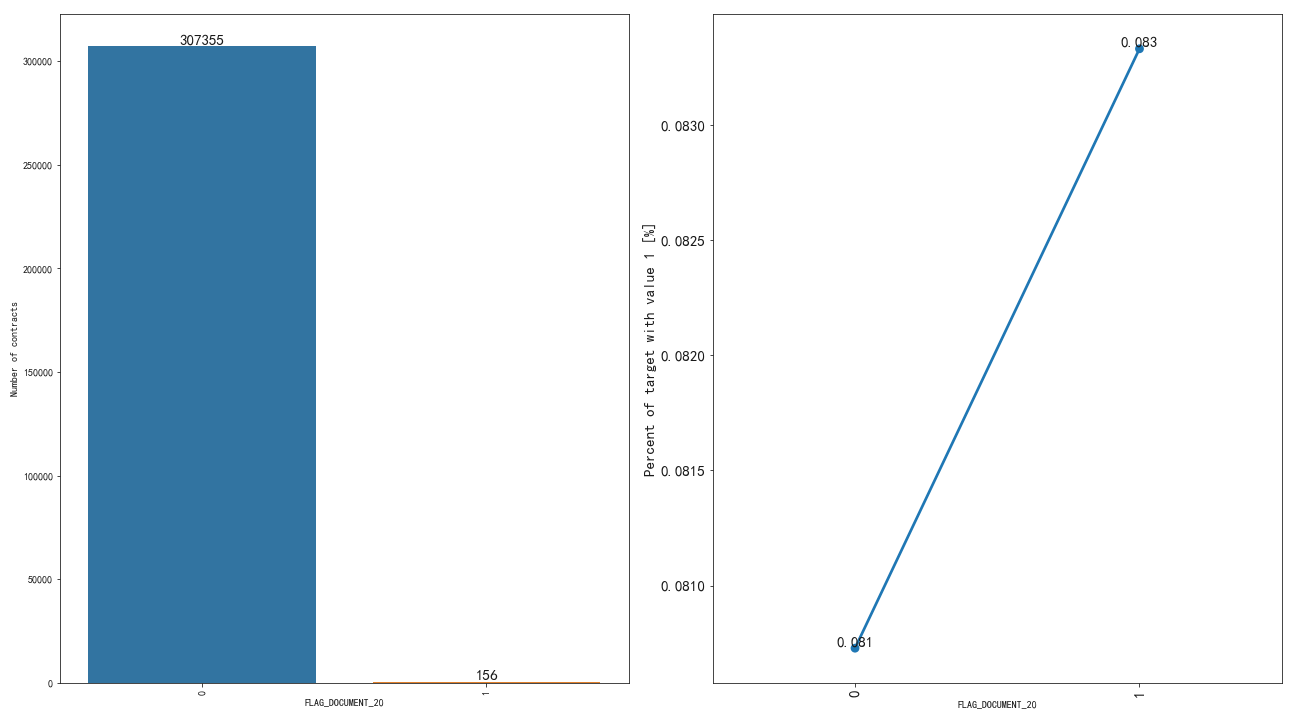












* 20个字段无缺失，为布尔型变量。
* 提供文件4-19的贷款记录占比非常非常小，flag=0的贷款记录的违约率接近于整体违约率，flag=1的违约率略小于整体违约率。
* 提供20文件的人也很少，flag=0与1时违约率很接近，可以说该字段对影响很小
* 原始特征包含的信息较少，可以提取特征后剔除
* 提供文件每条记录提供4至20的个数，再归一化？
* 提供2，21文件的人也很少，但是提供了这个文件却违约的贷款记录占比较高，违约率达到了0.3和0.13。远远高于整体违约率。
* 怎么处理？？
* 提供文件3的人较多，但提供了该文件的贷款记录违约率略高于为提供的。

## 电话邮件信息完整程度

