|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 变量名称 | 变量类型 | 变量的处理与赋值方式 | 变量对应数据表 | 备注 |
| **个人基本情况** | | | | |
| 是否本地籍 | 虚拟变量 | 0=非本地常住户口,1=本地常住户口 | 基础表 |  |
| 工作城市 | 分类变量 | 共284个城市 |  |
| 教育 | 顺序型变量 | 本科、博士研究生、初中、高中、其他、硕士及以上、硕士研究生、专科、专科及以下 | 分类有点乱，考虑分类合并 |
| 婚姻 | 分类变量 | 离婚、离异、其他、丧偶、未婚、已婚 |  |
| **个人经济状况** | | | | |
| 是否有公积金 | 虚拟变量 | 0=无， 1=有 | 基础表  基础表 |  |
| 收入 | 连续型变量 |  | 缺失严重，缺失约70% |
| **贷款情况** | | | | |
| 客户渠道 | 分类变量 | 共18种 | 基础表 | 缺失严重，缺失约70% |
| 账户状态 | 分类变量 | 呆账、结清、逾期、正常、转出 | 机构版征信-贷款 |  |
| 贷款种类细分 | 分类变量 | 9类贷款 | 机构版征信-信用提示表中统计了个人住房贷款笔数、个人商用房贷款笔数、其他贷款笔数 |
| 担保方式 | 分类变量 | 8种 |  |
| 还款频率 | 数值型变量 |  |  |
| 还款期数 | 数值型变量 |  |  |
| 五级分类 | 顺序型变量 | 次级、关注、可疑、未知、正常 | 缺失0.5305267 |
| 合同金额 | 数值型变量 |  |  |
| 本金余额 | 数值型变量 |  |  |
| 剩余还款期数 | 数值型变量 |  |  |
| 本月应还款 | 数值型变量 |  | 合并取差值 |
| 本月实还款 | 数值型变量 |  |
| 当前逾期期数 | 数值型变量 |  |  |
| 当前逾期金额 | 数值型变量 |  |  |
| 最近24个月还款状态 |  |  | 机构版征信-贷记卡 | 数据为文本符号。信息未知 |
| 个人住房贷款笔数 | 数值型变量 |  | 机构版征信-信用提示 | 与贷款种类细分变量关联 |
| 个人商用房（包括商住两用）贷款笔数 | 数值型变量 |  | 与贷款种类细分变量关联 |
| 其他贷款笔数 | 数值型变量 |  | 与贷款种类细分变量关联 |
| 贷记卡账户数 | 数值型变量 | 取值范围为1至3669 | 特征含义不明白，是否需要计算两种账户数比值作为新的特征 |
| 准贷记卡账户数 | 数值型变量 | 取值范围为0至15 |  |
| 贷款种类 | 分类数据 | 未结清贷款信息汇总、未销户贷记卡信息汇总、未销户准贷记卡信息汇总 | 机构版征信-未销户贷记卡或者未结清贷款 |  |
| 贷款法人数 | 数值型变量 |  |  |
| 贷款机构数 | 数值型变量 |  |  |
| 贷款账户数 | 数值型变量 |  |  |
| 合同金额 | 数值型变量 |  |  |
| 平均单个贷款机构最大合同金额 | 数值型变量 |  |  |
| 平均单个贷款机构最小合同金额 | 数值型变量 |  |  |
| 贷款余额 | 数值型变量 |  |  |
| 已用额度 | 数值型变量 |  |  |
| 最近6个月平均使用额度 | 数值型变量 |  |  |
| **信用情况** | | | | |
| 贷款种类 | 数值型变量 |  | 机构版征信-逾期(透支)信息汇总 |  |
| 贷款逾期笔数 | 数值型变量 |  |  |
| 贷款单月最高逾期总额 | 数值型变量 |  |  |
| 最大贷款时长 | 数值型变量 |  |  |
| 逾期/透支月数 | 数值型变量 |  |  |
| 逾期/透支金额 | 数值型变量 |  |  |
| 查询类别 | 分类变量 |  | 机构版征信-查询记录汇总 | 只有95各report\_id的数据  只有95各report\_id的数据 |
| 查询次数 | 数值型变量 |  |
| 查询机构 | 分类变量 | 4种 | 机构版征信-报告主表 |  |
| 查询原因 | 分类变量 | 3种 |  |

其他问题：时间指标涉及的特征不会分析。下表给出了竞赛数据各表名称以及对应的记录数、涉及的report\_id数，以及各表涉及的时间指标。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表格名称 | 记录数 | 涉及report\_id数 | 涉及的时间指标 | 备注 |
| contest\_basic\_train  （基础表-训练集）  contest\_basic\_test  （基础表-测试集） | 测试集10000条  训练集30000条 | 共40000 | 放款时间 |  |
| contest\_ext\_crd\_hd\_report  (机构版征信-报告主表) | 40000 | 40000 | 报告生成时间 |  |
| contest\_ext\_crd\_cd\_ln  （机构版征信-贷款） | 357196 | 35594 | 最近一次还款日期  应还款日  发放日期  到期日期 | 不明白这两个数据表的关系，两表变量基本相同，是否利用其中一表的信息即可。另外贷记卡表中币种变量有10种 |
| contest\_ext\_crd\_cd\_lnd  （机构版征信-贷记卡） | 324229 | 39850 | 发放日期  应还款日  最近一次还款日期 |
| contest\_ext\_crd\_is\_creditcue  （机构版征信-信用提示） | 39970 | 39970 |  |  |
| contest\_ext\_crd\_is\_sharedebt  （机构版征信-未销户贷记卡或者未结清贷款） | 76246 | 39948 |  |  |
| contest\_ext\_crd\_is\_ovdsummary  （机构版征信-逾期(透支)信息汇总） | 76212 | 25404 |  |  |
| contest\_ext\_crd\_qr\_recordsmr  （机构版征信-查询记录汇总） | 760 | 95 |  | 记录数量过少，是否不考虑该表信息 |
| contest\_ext\_crd\_qr\_recorddtlinfo  （机构版征信-信贷审批查询记录明细） | 654329 | 39865 |  |  |
| contest\_ext\_crd\_cd\_ln\_spl  （机构版征信-贷款特殊交易） | 67725 | 14190 |  |  |
| contest\_ext\_crd\_cd\_lnd\_ovd  （机构版征信-贷记卡逾期/透支记录） | 199644 | 16429 |  |  |