

目标变量：isDefault，ID变量：id

预处理前字段数：47，预处理后字段数：46

预处理前存在缺失值的字段数：22，预处理后存在缺失值的字段数：0

数据总行数：800,000，其中由于目标值缺失被删除行数：0

变量选出统计表			
	选出个数	未选出个数	总个数
全部变量	46	1	47
单值型变量	0	1	1
二值型变量	4	0	4
分类型变量	11	0	11
数值型变量	29	0	29
计数型变量	1	0	1
时间日期变量	0	0	0

变量处理信息

变量名称：id，变量类型：ID

变量名称：loanAmnt，变量类型：数值型变量

偏度：0； 平均值：14,416.819；

中位数：12,000； 方差：8,716.086。

该变量采用智能补缺算法填补缺失值。

变量名称：term，变量类型：二值型变量

分类个数：2

该变量采用智能补缺算法填补缺失值。

存在低频分类，2个分类由于频数过低被合并。

生成分类衍生变量：BI_term_1

变量名称：interestRate，变量类型：数值型变量

偏度：0； 平均值：13.238；

中位数：12.74； 方差：4.766。

该变量采用智能补缺算法填补缺失值。

变量名称：installment，变量类型：数值型变量

偏度：0； 平均值：437.948；

中位数：375.13； 方差：261.46。

该变量采用智能补缺算法填补缺失值。

变量名称：grade，变量类型：分类型变量

分类个数：7

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_grade

变量名称：subGrade，变量类型：分类型变量

分类个数：35

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_subGrade

变量名称：employmentTitle，变量类型：数值型变量

偏度：0； 平均值：72,005.352；

中位数：7,755； 方差：106,585.574。

该变量采用智能补缺算法填补缺失值。

变量名称：employmentLength，变量类型：分类型变量

分类个数：11

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_employmentLength

变量名称：homeOwnership，变量类型：分类型变量

分类个数：6

该变量采用智能补缺算法填补缺失值。

存在低频分类，6个分类由于频数过低被合并。

生成分类衍生变量：BI_homeOwnership_1, BI_homeOwnership_2,
BI_homeOwnership_3, BI_homeOwnership_4, BI_homeOwnership_5

变量名称：annualIncome，变量类型：数值型变量

偏度：0； 平均值：76,133.91；

中位数：65,000； 方差：68,947.514。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：verificationStatus，变量类型：分类型变量

分类个数：3

该变量采用智能补缺算法填补缺失值。

存在低频分类，3个分类由于频数过低被合并。

生成分类衍生变量：BI_verificationStatus_1, BI_verificationStatus_2

变量名称：issueDate，变量类型：分类型变量

分类个数：139

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_issueDate

变量名称：purpose，变量类型：分类型变量

分类个数：14

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_purpose

变量名称：postCode，变量类型：数值型变量

偏度：0； 平均值：258.536；

中位数：203； 方差：200.037。

该变量采用智能补缺算法填补缺失值。

变量名称：regionCode，变量类型：计数型变量

偏度：0； 平均值：16.386；

中位数：14； 方差：11.037。

该变量采用智能补缺算法填补缺失值。

变量名称：dti，变量类型：数值型变量

偏度：0； 平均值：18.285；

中位数：17.63； 方差：11.148。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：delinquency_2years，变量类型：数值型变量

偏度：0； 平均值：0.318；

中位数：0； 方差：0.88。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：ficoRangeLow，变量类型：数值型变量

偏度：0； 平均值：696.204；

中位数：690； 方差：31.866。

该变量采用智能补缺算法填补缺失值。

变量名称：ficoRangeHigh，变量类型：数值型变量

偏度：0； 平均值：700.204；

中位数：694； 方差：31.867。

该变量采用智能补缺算法填补缺失值。

变量名称：openAcc，变量类型：数值型变量

偏度：0； 平均值：11.598；

中位数：9； 方差：5.475。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：pubRec，变量类型：数值型变量

偏度：0； 平均值：0.215；

中位数：0； 方差：0.606。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：pubRecBankruptcies，变量类型：分类型变量

分类个数：11

该变量采用智能补缺算法填补缺失值。

存在低频分类，11个分类由于频数过低被合并。

生成分类衍生变量：BI_pubRecBankruptcies_1.0, BI_pubRecBankruptcies_2.0,
BI_pubRecBankruptcies_3.0, BI_pubRecBankruptcies_4.0

变量名称：revolBal，变量类型：数值型变量

偏度：0； 平均值：16,228.707；

中位数：11,132； 方差：22,458.021。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：revolUtil，变量类型：数值型变量

偏度：0； 平均值：51.791；

中位数：52.1； 方差：24.508。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：totalAcc，变量类型：数值型变量

偏度：0； 平均值：24.999；

中位数：23； 方差：11.999。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：initialListStatus，变量类型：二值型变量

分类个数：2

该变量采用智能补缺算法填补缺失值。

存在低频分类，2个分类由于频数过低被合并。

生成分类衍生变量：BI_initialListStatus_1

变量名称：applicationType，变量类型：二值型变量

分类个数：2

该变量采用智能补缺算法填补缺失值。

存在低频分类，2个分类由于频数过低被合并。

生成分类衍生变量：BI_applicationType_1

变量名称：earliesCreditLine，变量类型：分类型变量

分类个数：720

该变量采用智能补缺算法填补缺失值。

生成分类衍生变量：Smooth_earliesCreditLine

变量名称：title，变量类型：数值型变量

偏度：0； 平均值：1,754.114；

中位数：1； 方差：7,941.469。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n0，变量类型：数值型变量

偏度：0； 平均值：0.512；

中位数：0； 方差：1.299。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n1，变量类型：数值型变量

偏度：0； 平均值：3.642；

中位数：7； 方差：2.19。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n2，变量类型：数值型变量

偏度：0； 平均值：5.643；

中位数：4； 方差：3.219。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n3，变量类型：数值型变量

偏度：0； 平均值：5.643；

中位数：4； 方差：3.219。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n4，变量类型：数值型变量

偏度：0； 平均值：4.736；

中位数：3； 方差：2.888。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n5，变量类型：数值型变量

偏度：0； 平均值：8.108；

中位数：6； 方差：4.677。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n6，变量类型：数值型变量

偏度：0； 平均值：8.576；

中位数：6； 方差：7.212。

该变量采用智能补缺算法填补缺失值。

变量名称：n7，变量类型：数值型变量

偏度：0； 平均值：8.283；

中位数：8； 方差：4.445。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n8，变量类型：数值型变量

偏度：0； 平均值：14.622；

中位数：12； 方差：7.917。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n9，变量类型：数值型变量

偏度：0； 平均值：5.592；

中位数：5； 方差：3.134。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n10，变量类型：数值型变量

偏度：0； 平均值：11.644；

中位数：10； 方差：5.369。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n11，变量类型：分类型变量

分类个数：5

该变量采用智能补缺算法填补缺失值。

存在低频分类，5个分类由于频数过低被合并。

生成分类衍生变量：BI_n11_1.0, BI_n11_2.0, BI_n11_3.0, BI_n11_4.0

变量名称：n12，变量类型：分类型变量

分类个数：5

该变量采用智能补缺算法填补缺失值。

存在低频分类，5个分类由于频数过低被合并。

生成分类衍生变量：BI_n12_1.0, BI_n12_2.0, BI_n12_3.0, BI_n12_4.0

变量名称：n13，变量类型：数值型变量

偏度：0； 平均值：0.089；

中位数：0； 方差：0.496。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：n14，变量类型：数值型变量

偏度：0； 平均值：2.179；

中位数：2； 方差：1.797。

该变量采用智能补缺算法填补缺失值。

report.VClean

变量名称：isDefault，变量类型：二值型变量