



数据分析 项目介绍

➤ 流失率项目

➤ 信贷项目

流失率项目介绍

项目背景：根据调查，企业平均5年就会失去一半的客户，而失去一个客户带来的巨大损失，可能需要拉进来10个新客户才可以弥补。

项目意义：

- 挽回客户；
- 减小挽回工作的工作量。

流失定义：

- 抛出规则：统计日向前推5天，前3天有10笔以上大于2元的交易，后2天无交易的商户；
- 流失规则：抛出商户在后续28天中没有交易，类别划为流失。

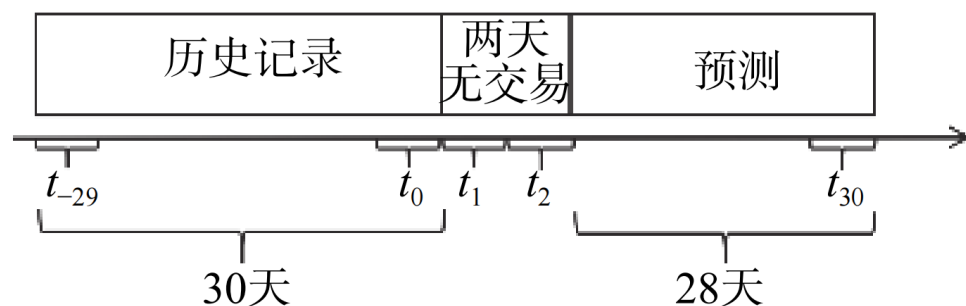


图 1 商户流失预测问题定义

项目目标：预测抛出商户中，哪些是未来28天内不再有交易（流失）的商户。

信贷项目介绍

项目背景:

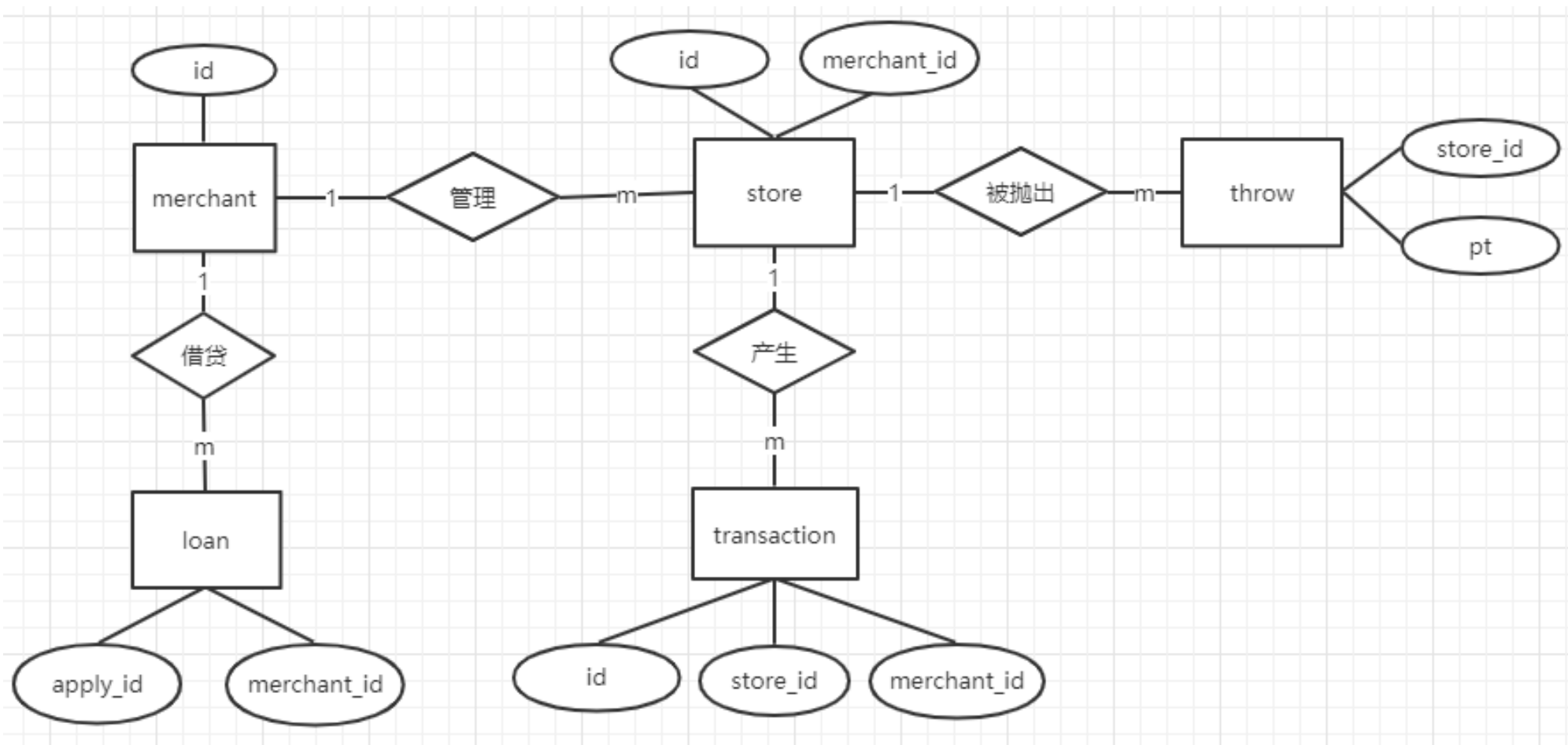
- 小微商户可以向平台贷款;
- 商户会存在逾期还款甚至不还款的行为。

项目意义: 减少贷款业务中出现坏账的可能性。

项目目标:

根据商户信息、信用记录和交易流水信息, 预测该条贷款申请是否会出现30天逾期还款现象。

实体关系



数据集介绍

- 2018年4月-2018年6月，商户的交易流水数据、贷款数据、还款数据等。

商户数据：

- table_datamining_merchant.txt

抛出点数据：

- table_throwpoint_training.txt
- table_throwpoint_test.txt

流水数据：

- table_loss_transaction_trainingdata.txt
- table_loss_transaction_testdata.txt
- table_lend_transaction_final.txt

历史还款数据：

- table_datamining_history_loan.txt

贷款数据：

- table_loan_train.txt
- table_loan_test.txt

表名：table_datamining_merchant

该表描述了商户的基本信息，包括：商户的行业，商户的省份。
商户的行业和省份在文件中都用数字表示。

列名	comment
merchant_id	商户ID
industry_level1	一级行业 [1,2,3]
province	省份 [1,2,3]

表名：table_throwpoint_training

表名：table_throwpoint_test

这两张表描述了被抛出店铺的ID和抛出的时间。

列名	comment
store_id	店铺ID
pt	抛出时间

列名	comment
store_id	店铺ID
pt	抛出时间

表名：table_loan_*

该表描述了贷款申请的详细信息以及标签。

列名	comment
apply_id	贷款申请id
merchant_id	商户贷款id
apply_date	贷款申请时间
lend_amount	贷款金额
lend_period	贷款期数
is_30days_overdue	有无逾期30天还款现象

表名：table_datamining_history_loan

该表描述了历史贷款申请的详细信息以及标签。

列名	comment
apply_id	贷款申请id
merchant_id	贷款人id
apply_date	贷款申请时间
lend_amount	贷款金额
lend_period	贷款期数
notional_settle_date	理论结清日
real_settle_date	实际结清日
is_30days_overdue	有无逾期30天还款现象

表名：table_transaction

该表描述了交易流水的详细信息。

列名	comment
id	交易id
type	交易类型 30-付款 10-取消 11-退款
status	状态 0-创建 2000-成功
store_id	门店id
merchant_id	商户id
pay_way	一级主支付方式
sub_pay_way	二级支付方式
paid_amount	消费者实际支付金额
ctime	创建时间
bankcard_credit	信用卡支付金额
bankcard_debit	储蓄卡支付金额
wallet_weixin	微信余额支付金额
wallet_alipay	支付宝余额支付金额
wallet_alipay_finance	支付宝余额宝支付金额
alipay_huabei	花呗支付金额
alipay_point	集分宝支付金额
pt	日期分区

数据和数据格式请勿放在网上