

现有一款现金分期产品，产品流程为 “用户注册-申请借款-风控审核-放款-用户还款-逾期催收”，假设数据存于 Mysql 数据库 risk 当中，现有用户信息 user 表、放款订单 loan 表和还款记录 repay 表，表结构如下：

user 表

字段名	use_id	phone_no	name	id_no	registered_from	registered_at
注释	用户编号	手机号	姓名	身份证号	注册渠道	注册时间
数据类型	bigint(20)	varchar(32)	varchar(50)	varchar(32)	bigint(20)	timestamp
缺省值	(主键)	null	null	null	null	CURRENT_TIMESTAMP
记录示例	100588	1883104*** *	张*	37584619900928*** *	556	2016-09-01 00:12:59

loan 表

loan_id	user_id	phone_no	applied_channel	applied_at	appraval_at	transacted_at	amount	term
订单编号	用户编号	手机号	申请渠道	申请时间	通过时间	放款时句	放款本金	分期数
111	100588	1883104*** *	222	2016-09-01 08:11:58	2016-09-01 08:12:54	2016-09-02 09:17:51	10000	6
...								

repay 表

id	loan_id	term no	prin	dead line	status	paid_prin	repaid_at
还款编号	订单编号	第几期	应还本金	到期日	还款状态	实还本金	实还时间
1234	111	1	3334	2016-10-02 23:59:59	3	3334	2016-10-02 19:16:50
...							

1. 请写 SQL 查出

(1) 注册渠道 556 上各放款月份的放款件数和首期到期 0 天的件数逾期率。注：首期即还款期数的第一期，0 天件数逾期率=逾期 1 天及以上的件数/所有到期件数

2. 现有如下表头的 Excel 文件 model3.xlsx，请用 python/R 读取 excel 存入变量 df 中，对 loan_id 做去重处理，并输出各放款月(transacted_at) 各分期数(term) 的样本量分布。

loan_id	transacted at	amount	term	content	model3
23512	2017-07-09 20:34:56	10000	9	{ "loan_id ":"23512", "data":{"A":1, "B":2}}	0.1257465

3. 请用 python/R 在变量 df 中新增字段 A 和B，存入 content 中对应的 json 数据，并处理缺失值。

以下两题任选一题作答即可

4. 变量 df 中 transacted_at 在 18-05-01 至 18-07-01 的样本为模型 model3 的测试集，请用 python/R 筛选出测试集中 9 期产品 (term=9) 的样本，并画出模型 model3 在该样本上 ROC 曲线和 Lift Chart 提升图。

5. 以下是该现金分期产品的额度策略：分为初始额度和每次提额。

根据客户质量分为三组，初始额度由高至低为500元，400元，300元。

每次复贷提额策略分别为上一次额度乘以一个系数，再减去一个常量。然后四舍五入取整。

例：优质客户第一次借款，额度为500元，第二次借款时额度为 round (500 + (500*25% - 25))。

现有一客户额度是2006元，请用**递归函数**求出其所在分组和提额次数（初始额度提额次数为0）。

	优质客户	中等客户	低质客户
初始额度	¥500	¥400	¥300
每次提额	25% - ¥25	20% - ¥20	15% - ¥15