

数据分析场景介绍

场景1: 30天+逾期率分析

业务背景

逾期率分析是金融风控的核心指标之一，通过分析不同放款月份在不同账龄（MOB）下的逾期表现，帮助评估资产质量和制定风控策略。

分析目标

生成逾期率分析表，纵坐标为放款月份，横坐标为MOB（Month of Book，账龄月份），数值为30天+逾期率（逾期金额/总放款金额）。

数据源定义

表1: loan_info (放款信息表)

```
CREATE TABLE loan_info (  
  loan_id VARCHAR(50) PRIMARY KEY,      -- 放款ID  
  loan_amount DECIMAL(15,2),             -- 放款金额  
  loan_date DATE,                        -- 放款日期  
  loan_month VARCHAR(7),                 -- 放款月份 (YYYY-MM)  
  borrower_id VARCHAR(50),               -- 借款人ID  
  product_type VARCHAR(20),              -- 产品类型  
  loan_term INT,                         -- 放款期限(月)  
  interest_rate DECIMAL(5,4),            -- 利率  
  status VARCHAR(20) DEFAULT 'ACTIVE'    -- 放款状态  
);
```

表2: lending_details (还款明细表)

```
CREATE TABLE lending_details (  
    detail_id VARCHAR(50) PRIMARY KEY,          -- 明细ID  
    loan_id VARCHAR(50),                        -- 放款ID (外键)  
    due_date DATE,                             -- 应还日期  
    actual_payment_date DATE,                  -- 实际还款日期  
    due_amount DECIMAL(15,2),                  -- 应还金额  
    actual_amount DECIMAL(15,2),               -- 实际还款金额  
    overdue_days INT DEFAULT 0,                -- 逾期天数  
    overdue_amount DECIMAL(15,2) DEFAULT 0,    -- 逾期金额  
    payment_status VARCHAR(20),               -- 还款状态 (PAID/OVERDUE/PENDING)  
    mob INT,                                   -- 账龄月份 (Month of Book)  
    created_date TIMESTAMP DEFAULT CURRENT_TIMESTAMP,  
    FOREIGN KEY (loan_id) REFERENCES loan_info(loan_id)  
);
```

表3: customer_info (客户个人信息表)

```
CREATE TABLE customer_info (  
    customer_id VARCHAR(50) PRIMARY KEY,        -- 客户ID  
    customer_name VARCHAR(100),                 -- 客户姓名  
    id_card VARCHAR(18),                        -- 身份证号  
    phone VARCHAR(20),                          -- 手机号  
    gender VARCHAR(10),                        -- 性别  
    age INT,                                    -- 年龄  
    education VARCHAR(20),                     -- 学历  
    marital_status VARCHAR(20),                -- 婚姻状况  
    occupation VARCHAR(50),                   -- 职业  
    monthly_income DECIMAL(15,2),              -- 月收入  
    annual_income DECIMAL(15,2),              -- 年收入  
    credit_score INT,                          -- 信用评分  
    province VARCHAR(50),                     -- 省份  
    city VARCHAR(50),                          -- 城市  
    district VARCHAR(50),                     -- 区县  
    debt_to_income_ratio DECIMAL(5,4),         -- 负债收入比  
    employment_years INT,                     -- 工作年限  
    house_ownership VARCHAR(20),              -- 房产状况  
    car_ownership VARCHAR(20),                -- 车产状况  
    created_time TIMESTAMP DEFAULT CURRENT_TIMESTAMP  
);
```

关键字段说明

- **loan_month**: 放款月份，格式为 YYYY-MM
- **mob**: 账龄月份，表示放款后第几个月
- **overdue_days**: 逾期天数，≥30天视为逾期
- **overdue_amount**: 逾期金额
- **loan_amount**: 总放款金额

分析需求

输入数据格式

- **数据源**: lending_details表 + loan_info表 (通过loan_id关联)
- **关键字段**:
 - loan_month : 放款月份 (YYYY-MM格式)
 - mob : 账龄月份 (1, 2, 3, 6, 12, 24等)
 - overdue_days : 逾期天数 (≥30天视为逾期)
 - overdue_amount : 逾期金额
 - loan_amount : 放款金额

计算逻辑

1. **逾期定义**: overdue_days ≥ 30天
2. **逾期率计算公式**:

$$\text{逾期率} = \frac{\sum(\text{30天+逾期金额})}{\sum(\text{对应MOB的总放款金额})} \times 100\%$$

3. **分组维度**: 按放款月份(loan_month) × MOB进行交叉分析

预期输出格式

生成vintage分析交叉表，行为放款月份，列为不同MOB，值为逾期率百分比:

放款月份	MOB1	MOB2	MOB3	MOB6	MOB12	MOB24
2023-01	0.5%	1.2%	2.1%	3.8%	5.2%	6.1%
2023-02	0.4%	1.1%	2.0%	3.5%	4.9%	-
2023-03	0.6%	1.3%	2.2%	3.9%	-	-

注: "-" 表示该MOB尚未到期或数据不足

预期输出格式

放款月份	MOB1	MOB2	MOB3	MOB6	MOB12	MOB24
2023-01	0.5%	1.2%	2.1%	3.8%	5.2%	6.1%
2023-02	0.4%	1.1%	2.0%	3.5%	4.9%	-
2023-03	0.6%	1.3%	2.2%	3.9%	-	-
...						

场景2: 逾期根因分析

分析目标

基于逾期率矩阵输出结果，深入分析导致逾期的根本原因，识别高风险因素，为风控策略优化提供数据支持。

根因分析维度

1. 利率因素分析

分析假设: 利率水平直接影响借款人还款压力，高利率可能导致逾期率上升

关键发现示例:

- 利率≤24%时，MOB6逾期率约为3.2%
- 利率在24%-35%区间时，MOB6逾期率上升至5.8%
- 利率>35%时，MOB6逾期率激增至12.5%
- 临界点: 35%为逾期率急剧上升的临界利率

业务建议:

- 对于利率>35%的产品，需要加强风控审核
- 考虑设置利率上限，平衡收益与风险
- 针对高利率客户提供还款提醒和分期服务

2. 贷款金额因素分析

分析假设: 大额贷款增加借款人还款负担，可能提高逾期风险

关键发现示例:

- 贷款金额≤5,000元时，MOB6逾期率为2.1%

- 贷款金额5,000-10,000元时，MOB6逾期率为4.3%
- 贷款金额>10,000元时，MOB6逾期率上升至8.7%
- **临界点:** 10,000元为逾期率显著上升的临界金额

业务建议:

- 对于>10,000元的大额贷款，提高审核标准
- 增加收入证明和资产证明要求
- 考虑分期放款或设置额度上限

3. 地区因素分析

分析假设: 不同地区的经济发展水平、就业环境影响还款能力

关键发现示例:

- **高风险省份 (MOB6逾期率>8%):**
 - 青海省: 12.3%
 - 甘肃省: 10.8%
 - 新疆维吾尔自治区: 9.5%
- **中风险省份 (MOB6逾期率4%-8%):**
 - 河南省: 6.7%
 - 安徽省: 5.9%
 - 湖南省: 5.2%
- **低风险省份 (MOB6逾期率<4%):**
 - 北京市: 2.1%
 - 上海市: 2.3%
 - 广东省: 3.2%

业务建议:

- 对高风险省份实施更严格的风控政策
- 调整不同地区的利率定价策略
- 加强高风险地区的贷后管理

4. 个人信息因素分析

4.1 年龄因素

关键发现示例:

- 18-25岁: MOB6逾期率7.8% (缺乏信用记录, 收入不稳定)

- 26-35岁: MOB6逾期率4.2% (收入稳定期)
- 36-45岁: MOB6逾期率3.1% (事业稳定期)
- 46-55岁: MOB6逾期率5.6% (面临中年危机)
- 55岁以上: MOB6逾期率8.9% (收入下降, 健康风险)

4.2 收入因素

关键发现示例:

- 月收入<3,000元: MOB6逾期率11.2%
- 月收入3,000-8,000元: MOB6逾期率6.8%
- 月收入8,000-15,000元: MOB6逾期率3.9%
- 月收入>15,000元: MOB6逾期率2.1%

4.3 学历因素

关键发现示例:

- 初中及以下: MOB6逾期率9.3%
- 高中/中专: MOB6逾期率6.7%
- 大专: MOB6逾期率4.8%
- 本科: MOB6逾期率3.2%
- 硕士及以上: MOB6逾期率1.9%

4.4 职业因素

关键发现示例:

- **高风险职业:**
 - 个体户: 10.5%
 - 服务员: 9.8%
 - 销售员: 8.7%
- **中风险职业:**
 - 工人: 6.2%
 - 司机: 5.9%
- **低风险职业:**
 - 公务员: 1.8%
 - 教师: 2.3%
 - 医生: 2.7%

4.5 婚姻状况因素

关键发现示例:

- 未婚: MOB6逾期率6.8%
- 已婚: MOB6逾期率3.9%
- 离异: MOB6逾期率8.2%
- 丧偶: MOB6逾期率7.1%

4.6 信用评分因素

关键发现示例:

- 信用评分<500: MOB6逾期率15.6%
- 信用评分500-600: MOB6逾期率8.9%
- 信用评分600-700: MOB6逾期率5.2%
- 信用评分700-800: MOB6逾期率2.8%
- 信用评分>800: MOB6逾期率1.1%

5. 组合因素分析

5.1 高风险组合识别

极高风险组合 (MOB6逾期率>20%):

- 年龄18-25岁 + 月收入<3,000元 + 利率>35% + 贷款金额>10,000元
- 个体户 + 初中学历 + 离异 + 高风险省份

高风险组合 (MOB6逾期率10%-20%):

- 年龄55岁以上 + 月收入<5,000元 + 利率>30%
- 服务员 + 未婚 + 贷款金额>8,000元

5.2 低风险组合识别

极低风险组合 (MOB6逾期率<2%):

- 公务员 + 本科以上学历 + 已婚 + 月收入>10,000元 + 利率<24%
- 医生/教师 + 信用评分>750 + 低风险省份

根因分析输出格式

1. 单因素风险排序表

风险因素	因素取值	MOB6逾期率	风险等级	样本量
利率	>35%	12.5%	极高风险	1,250
信用评分	<500	15.6%	极高风险	890
职业	个体户	10.5%	高风险	2,100
省份	青海省	12.3%	极高风险	450
年龄	18-25岁	7.8%	中高风险	3,200

2. 多因素交叉分析表

年龄组	收入组	利率区间	逾期率	风险等级	建议措施
18-25岁	<3K	>35%	25.6%	极高风险	拒绝放款
26-35岁	3K-8K	24%-35%	8.9%	中风险	加强审核
36-45岁	>8K	<24%	2.1%	低风险	正常放款

3. 地区风险热力图数据

省份	MOB3逾期率	MOB6逾期率	MOB12逾期率	风险等级
北京市	1.2%	2.1%	3.8%	低风险
上海市	1.1%	2.3%	4.1%	低风险
广东省	1.8%	3.2%	5.6%	低风险
河南省	3.2%	6.7%	11.2%	中风险
青海省	6.8%	12.3%	18.9%	极高风险

业务价值

- **风险监控:** 及时发现资产质量恶化趋势
- **策略调整:** 基于不同vintage表现调整风控策略
- **预算规划:** 预估未来损失准备金需求
- **产品优化:** 识别高风险产品和客群
- **精准定价:** 基于风险因素调整利率定价
- **客群细分:** 识别优质客群和高风险客群
- **政策制定:** 为不同地区制定差异化风控政策
- **预警机制:** 建立多维度风险预警体系

场景3: 静态逾期率单因素分析

业务背景

静态逾期率分析是风险管理的重要工具，通过计算剩余本金与初始放款本金的比例，评估不同因素对逾期风险的影响程度。与动态逾期率不同，静态逾期率反映的是某一时点的资产质量状况。

分析目标

基于静态逾期率（剩余本金/初始放款本金）进行单因素分析，识别各个风险因素的影响程度，为风控策略制定提供量化依据。

静态逾期率定义

静态逾期率 = 剩余本金 / 初始放款本金

数据源扩展

表结构补充: lending_details (还款明细表)

```
ALTER TABLE lending_details ADD COLUMN (  
    remaining_principal DECIMAL(15,2),      -- 剩余本金  
    initial_principal DECIMAL(15,2),        -- 初始本金  
    dpd_days INT DEFAULT 0,                 -- DPD天数 (Days Past Due)  
    static_overdue_rate DECIMAL(8,6)        -- 静态逾期率  
);
```

单因素分析维度

1. 年龄因素静态逾期率分析

分析目标: 评估不同年龄段的静态逾期率表现

预期输出格式:

年龄段	样本量	平均静态逾期率	风险等级	风险排序
18-25岁	3,200	0.1580	高风险	1
26-35岁	8,500	0.0890	中风险	3
36-45岁	6,800	0.0650	低风险	5
46-55岁	4,200	0.1120	中高风险	2
55岁以上	2,100	0.1350	高风险	1

2. 收入因素静态逾期率分析

分析目标: 评估不同收入水平的静态逾期率表现

预期输出格式:

月收入区间	样本量	平均静态逾期率	标准差	风险等级
<3,000元	2,800	0.2150	0.185	极高风险
3,000-5,000元	4,500	0.1420	0.156	高风险
5,000-8,000元	6,200	0.0980	0.124	中风险
8,000-15,000元	5,800	0.0620	0.089	低风险
>15,000元	3,200	0.0380	0.067	极低风险

3. 地区因素静态逾期率分析

分析目标: 识别高风险地区，制定差异化风控策略

预期输出格式:

省份	样本量	平均静态逾期率	中位数	90分位数	风险等级
青海省	450	0.2890	0.245	0.456	极高风险
甘肃省	680	0.2340	0.198	0.398	极高风险
新疆维吾尔	520	0.2180	0.189	0.387	高风险
河南省	2,100	0.1560	0.134	0.289	中高风险
安徽省	1,800	0.1420	0.125	0.267	中风险
北京市	1,200	0.0450	0.038	0.089	极低风险
上海市	1,100	0.0480	0.041	0.092	极低风险
广东省	2,800	0.0720	0.062	0.134	低风险

4. 职业因素静态逾期率分析

分析目标: 识别高风险职业类别

预期输出格式:

职业类别	样本量	平均静态逾期率	风险系数	风险等级	建议措施
个体户	2,100	0.2680	2.45	极高风险	严格审核
服务员	1,800	0.2340	2.14	极高风险	提高准入门槛
销售员	2,400	0.1980	1.81	高风险	加强贷后管理
工人	3,200	0.1450	1.33	中风险	标准审核
司机	1,600	0.1380	1.26	中风险	标准审核
公务员	1,200	0.0420	0.38	极低风险	优先客户
教师	980	0.0480	0.44	极低风险	优先客户
医生	750	0.0520	0.48	极低风险	优先客户

5. 产品类型因素静态逾期率分析

分析目标: 评估不同产品的风险表现

预期输出格式:

产品类型	样本量	平均静态逾期率	利润率	风险调整收益	建议
消费贷	8,500	0.1280	12.5%	8.2%	优化定价
经营贷	4,200	0.1650	15.8%	9.1%	加强审核
装修贷	2,800	0.1120	10.2%	7.8%	标准产品
教育贷	1,900	0.0890	8.5%	6.9%	低风险产品
车贷	3,600	0.0950	9.8%	7.5%	标准产品

6. 信用评分因素静态逾期率分析

分析目标: 验证信用评分的有效性

预期输出格式:

信用评分区间	样本量	平均静态逾期率	违约概率	KS值	模型有效性
<500	890	0.3450	34.5%	0.68	强区分度
500-550	1,200	0.2780	27.8%	0.65	强区分度
550-600	1,800	0.2150	21.5%	0.62	强区分度
600-650	2,400	0.1620	16.2%	0.58	中等区分度
650-700	2,800	0.1180	11.8%	0.54	中等区分度
700-750	2,200	0.0780	7.8%	0.49	中等区分度
750-800	1,600	0.0450	4.5%	0.43	弱区分度
>800	1,100	0.0280	2.8%	0.38	弱区分度

单因素分析输出报告

1. 风险因素重要性排序

排序	风险因素	影响系数	解释度(R ²)	业务重要性
1	信用评分	0.456	0.312	极高
2	月收入	0.389	0.267	极高
3	职业类别	0.334	0.198	高
4	年龄	0.298	0.156	高
5	地区	0.267	0.134	中高
6	产品类型	0.189	0.089	中

2. 单因素风险预警阈值

风险因素	预警阈值	建议措施
信用评分	<550	拒绝放款
月收入	<3,000元	严格审核
职业	个体户/服务员	提高利率
年龄	18-25岁或>55岁	加强审核
地区	青海/甘肃/新疆	区域限制
产品类型	经营贷	降低额度

场景4: 基于DPD天数的多维度分析

业务背景

DPD（Days Past Due）分析是金融机构风险管理的核心工具，通过分析不同逾期天数阈值下的客户表现，实现精细化的风险识别和管理。

分析目标

基于DPD天数进行多维度分析，包括1+、7+、30+、60+、90+天等不同阈值，全面评估资产质量和客户风险。

DPD分析维度定义

1. DPD阈值分类

- DPD1+：逾期1天及以上
- DPD7+：逾期7天及以上
- DPD30+：逾期30天及以上
- DPD60+：逾期60天及以上
- DPD90+：逾期90天及以上
- DPD180+：逾期180天及以上

2. 数据源扩展

lending_details表DPD字段补充

```
ALTER TABLE lending_details ADD COLUMN (  
    dpd_base INT DEFAULT 0, -- DPD基准天数  
    dpd_1_plus BOOLEAN DEFAULT FALSE, -- DPD1+标识  
    dpd_7_plus BOOLEAN DEFAULT FALSE, -- DPD7+标识  
    dpd_30_plus BOOLEAN DEFAULT FALSE, -- DPD30+标识  
    dpd_60_plus BOOLEAN DEFAULT FALSE, -- DPD60+标识  
    dpd_90_plus BOOLEAN DEFAULT FALSE, -- DPD90+标识  
    dpd_180_plus BOOLEAN DEFAULT FALSE, -- DPD180+标识  
    max_dpd_ever INT DEFAULT 0 -- 历史最大DPD天数  
);
```

DPD多维度分析框架

1. DPD迁移率分析

分析目标: 观察不同DPD阶段之间的迁移规律

预期输出格式:

起始状态	目标状态	迁移率	样本量	置信区间
正常	DPD1+	8.5%	10,000	[8.1%, 8.9%]
DPD1+	DPD7+	35.2%	850	[32.1%, 38.3%]
DPD7+	DPD30+	45.8%	299	[40.2%, 51.4%]
DPD30+	DPD60+	52.3%	137	[44.1%, 60.5%]
DPD60+	DPD90+	68.9%	72	[57.3%, 80.5%]
DPD90+	DPD180+	78.4%	50	[65.2%, 91.6%]

2. 不同DPD阈值下的客户分布

分析目标: 了解各DPD阶段的客户规模和占比

预期输出格式:

DPD阶段	客户数量	占总客户比例	逾期金额(万元)	占总金额比例
正常	45,230	91.5%	0	0%
DPD1+	2,180	4.4%	1,250	2.8%
DPD7+	1,120	2.3%	980	2.2%
DPD30+	650	1.3%	890	2.0%
DPD60+	180	0.4%	420	0.9%
DPD90+	90	0.2%	380	0.8%
DPD180+	50	0.1%	320	0.7%

3. 基于DPD的客户画像分析

3.1 年龄维度DPD分析

年龄段	DPD1+率	DPD7+率	DPD30+率	DPD60+率	DPD90+率
18-25岁	12.8%	6.9%	4.2%	1.8%	1.2%
26-35岁	7.2%	3.8%	2.1%	0.9%	0.6%
36-45岁	5.9%	2.9%	1.6%	0.7%	0.4%
46-55岁	8.4%	4.5%	2.8%	1.3%	0.9%
55岁以上	11.6%	6.8%	4.5%	2.1%	1.5%

3.2 收入维度DPD分析

月收入区间	DPD1+率	DPD7+率	DPD30+率	DPD60+率	DPD90+率
<3,000元	18.9%	11.2%	7.8%	4.2%	3.1%
3,000-5,000元	12.4%	7.1%	4.5%	2.1%	1.6%
5,000-8,000元	8.7%	4.6%	2.8%	1.2%	0.8%
8,000-15,000元	5.2%	2.4%	1.3%	0.6%	0.3%
>15,000元	3.1%	1.2%	0.7%	0.2%	0.1%

3.3 地区维度DPD分析

省份	DPD1+率	DPD7+率	DPD30+率	DPD60+率	DPD90+率	风险等级
青海省	21.5%	13.8%	9.2%	5.1%	3.8%	极高风险
甘肃省	18.9%	11.4%	7.8%	4.2%	3.1%	极高风险
新疆维吾尔	17.2%	10.1%	6.9%	3.6%	2.7%	高风险
河南省	11.8%	6.7%	4.1%	1.9%	1.3%	中高风险
安徽省	10.2%	5.8%	3.5%	1.6%	1.1%	中风险
北京市	3.2%	1.4%	0.8%	0.3%	0.2%	极低风险
上海市	3.5%	1.6%	0.9%	0.4%	0.2%	极低风险
广东省	5.1%	2.3%	1.4%	0.6%	0.4%	低风险

4. DPD预警体系

4.1 早期预警指标

预警级别	DPD阈值	触发条件	预警措施
绿色	DPD0	正常还款	常规管理
黄色	DPD1-6	连续逾期1-6天	电话提醒
橙色	DPD7-29	连续逾期7-29天	上门催收
红色	DPD30-89	连续逾期30-89天	法务介入
黑色	DPD90+	连续逾期90天以上	资产处置

4.2 DPD风险评分模型

-- DPD风险评分计算公式

DPD风险评分 =

```
CASE
  WHEN dpd_base = 0 THEN 100
  WHEN dpd_base BETWEEN 1 AND 6 THEN 85 - dpd_base * 2
  WHEN dpd_base BETWEEN 7 AND 29 THEN 70 - (dpd_base - 6) * 1.5
  WHEN dpd_base BETWEEN 30 AND 89 THEN 35 - (dpd_base - 29) * 0.5
  WHEN dpd_base >= 90 THEN 5
END
```

5. DPD趋势分析

5.1 月度DPD趋势监控

月份	DPD1+率	DPD7+率	DPD30+率	环比变化	同比变化
2024-01	8.2%	4.1%	2.3%	-	-
2024-02	8.7%	4.5%	2.6%	↑5.8%	-
2024-03	9.1%	4.8%	2.9%	↑4.6%	-
2024-04	8.9%	4.6%	2.7%	↓2.2%	-
2024-05	8.5%	4.3%	2.4%	↓4.5%	-
2024-06	8.8%	4.7%	2.8%	↑3.5%	-

5.2 DPD季节性分析

季度	平均DPD1+率	平均DPD30+率	季节性因子	主要原因
Q1	9.2%	2.8%	1.15	春节影响
Q2	8.1%	2.4%	1.01	收入稳定
Q3	7.8%	2.2%	0.97	收入增长
Q4	8.9%	2.7%	1.11	年底支出增加

DPD分析业务价值

1. 风险管理价值

- 早期识别: 通过DPD1+快速识别潜在风险客户
- 精准干预: 基于不同DPD阶段制定差异化催收策略
- 损失预测: 通过DPD迁移率预测未来损失
- 资源配置: 基于DPD分布优化催收资源配置

2. 业务决策支持

- 产品定价: 基于DPD表现调整不同客群的利率
- 额度管理: 根据DPD历史调整客户授信额度
- 准入策略: 基于DPD分析优化客户准入标准
- 区域策略: 制定差异化的区域风控政策

3. 监管合规

- 监管报告: 满足监管机构对DPD数据的报告要求
- 压力测试: 基于DPD数据进行压力测试
- 风险计量: 支持巴塞尔协议下的风险计量

- **内控管理:** 建立基于DPD的内控管理体系