

AI-SaaS 工具留存与转化漏斗优化方案及业务复盘报告

项目名称: AI-SaaS 工具留存与转化漏斗优化项目
负责人: 数据产品运营
项目周期: 2026.07 — 2026.08
输出日期: 2026 年 08 月 16 日
项目来源: 和鲸社区 (ModelWhale)
项目链接: <https://www.heywhale.com/mw/project/6a8169684e551f2d69b5609f?shareby=6a81643b24ec58806f3f27df#>
作品集链接: <https://smartflow-ui.pages.dev>

一、项目背景

AI 工具 SaaS 行业普遍面临三大增长瓶颈:

- 新用户激活率偏低:** 注册后未完成关键动作的用户占比居高不下, 大量用户在首次体验前即流失。
- 核心功能渗透率不足:** 即便完成激活, 用户对平台核心价值功能的采用率仍低于预期, 导致产品粘性薄弱。
- 免费转付费链路损耗严重:** 从核心功能体验到首次付费的转化率存在显著断崖, 7 日新用户流失居高不下。

业务侧虽已具备较完善的埋点数据采集能力, 但缺少系统化的漏斗归因分析与用户分层洞察, 产品迭代与运营决策缺乏数据支撑。

本项目旨在通过 SQL 驱动的数据分析, 构建完整的转化漏斗模型, 定位流失卡点, 输出可落地的产品与运营策略。

二、数据基础说明

2.1 数据来源与性质

重要声明: 本项目数据集为使用 Python (Faker + NumPy + Pandas) 程序化生成的**模拟数据**, 非真实业务埋点数据。

维度	说明
生成工具	Python Faker + NumPy + Pandas
用户规模	5,000 名模拟用户
数据表数量	5 张 (users / events / sessions / subscriptions / saas_user_funnel)
时间范围	模拟 90 天行为周期

概率分布	参照真实 SaaS 业务常见分布设置 (如免费版 65%、专业版 25%、企业版 10%)
数据性质	模拟生成, 不涉及真实用户隐私

2.2 数据表结构

users 表 (用户主表)

字段	类型	说明
user_id	VARCHAR	用户唯一标识 (主键, 如 U100001)
signup_time	DATETIME	注册时间
plan	VARCHAR	订阅计划 (free / pro / enterprise)
industry	VARCHAR	所属行业
company_size	VARCHAR	公司规模 (1-10 / 11-50 / 51-200 / 200+)
mrr	INT	月经常性收入 (Monthly Recurring Revenue)
is_activate	INT	是否激活 (0/1)
is_pay	INT	是否付费 (0/1)
retain_7d	INT	是否 7 日留存 (0/1)
guide_complete_rate	FLOAT	新手引导完成度 (0.0 — 1.0)
core_used	INT	是否使用核心功能 (0/1)
channel	VARCHAR	获客渠道

events 表 (行为事件表)

字段	类型	说明
event_id	VARCHAR	事件唯一标识 (主键)
user_id	VARCHAR	外键, 关联 users 表
event_name	VARCHAR	事件名称 (login / view_dashboard / use_core_feature / upgrade / churn 等)
event_time	DATETIME	事件发生时间
duration_sec	INT	事件持续时长 (秒)

sessions 表 (会话表)

字段	类型	说明
----	----	----

session_id	VARCHAR	会话唯一标识（主键）
user_id	VARCHAR	外键，关联 users 表
start_time	DATETIME	会话开始时间
end_time	DATETIME	会话结束时间
page_views	INT	页面浏览数

subscriptions 表（订阅表）

字段	类型	说明
sub_id	VARCHAR	订阅唯一标识（主键）
user_id	VARCHAR	外键，关联 users 表
plan	VARCHAR	订阅计划
mrr	INT	月经常性收入
start_date	DATE	订阅开始日期
status	VARCHAR	订阅状态（active / cancelled / expired）

saas_user_funnel 表（漏斗标志表）

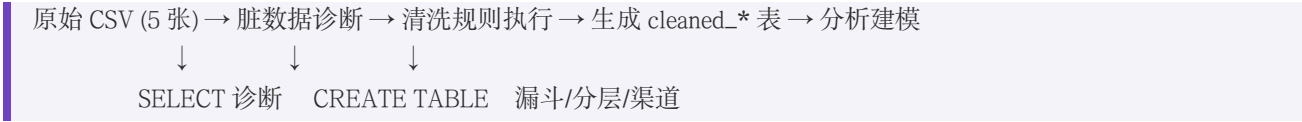
字段	类型	说明
user_id	VARCHAR	用户唯一标识
is_activate	INT	是否激活
core_used	INT	是否使用核心功能
is_pay	INT	是否付费
retain_7d	INT	是否 7 日留存
guide_complete_rate	FLOAT	新手引导完成度
channel	VARCHAR	获客渠道

2.3 数据价值定位

本项目的核心价值在于展示从数据清洗到漏斗建模、用户分层、渠道分析、策略落地的完整方法论，而非具体数值的绝对准确性。在真实业务场景中，相同的方法论与 SQL 逻辑可直接应用于实际埋点数据。

三、SQL 数据清洗方法论

3.1 清洗流程总览



3.2 第 1 步：脏数据诊断（先查再洗）

在执行清洗前，先全面诊断数据问题，避免盲目清洗导致信息丢失。

```
-- ===== 1.1 重复主键检查 =====
-- users 表
SELECT user_id, COUNT(*) AS cnt
FROM users
GROUP BY user_id
HAVING cnt > 1;

-- events 表
SELECT event_id, COUNT(*) AS cnt
FROM events
GROUP BY event_id
HAVING cnt > 1;

-- sessions 表
SELECT session_id, COUNT(*) AS cnt
FROM sessions
GROUP BY session_id
HAVING cnt > 1;

-- ===== 1.2 空值检查 =====
SELECT
  SUM(CASE WHEN user_id IS NULL THEN 1 ELSE 0 END) AS null_uid,
  SUM(CASE WHEN plan IS NULL THEN 1 ELSE 0 END) AS null_plan,
  SUM(CASE WHEN industry IS NULL THEN 1 ELSE 0 END) AS null_industry,
  SUM(CASE WHEN mrr IS NULL THEN 1 ELSE 0 END) AS null_mrr
FROM users;

-- ===== 1.3 异常值检查 =====
SELECT
  MIN(duration_sec) AS min_dur,
  MAX(duration_sec) AS max_dur,
  MIN(mrr) AS min_mrr,
  MAX(mrr) AS max_mrr
FROM events, subscriptions;
```

```
-- ===== 1.4 外键完整性检查 (孤儿记录) =====
SELECT COUNT(*) AS orphan_events
FROM events e
LEFT JOIN users u ON e.user_id = u.user_id
WHERE u.user_id IS NULL;
```

3.3 第 2 步：去重处理

兼容性说明：和鲸云端 SQLite 版本 < 3.25，不支持 `ROW_NUMBER() OVER()` 窗口函数，需用子查询替代实现等价去重。

```
-- ===== 方案 A: MySQL (支持窗口函数) =====
CREATE TABLE cleaned_users AS
SELECT * FROM (
  SELECT *,
    ROW_NUMBER() OVER(PARTITION BY user_id ORDER BY signup_time) AS rn
  FROM users
)t WHERE rn = 1;
```

```
-- ===== 方案 B: SQLite 兼容写法 (子查询去重) =====
CREATE TABLE cleaned_users AS
SELECT u1.*
FROM users u1
WHERE u1.rowid = (
  SELECT MIN(u2.rowid)
  FROM users u2
  WHERE u2.user_id = u1.user_id
);
```

```
-- events 表去重
CREATE TABLE cleaned_events AS
SELECT e1.*
FROM events e1
WHERE e1.rowid = (
  SELECT MIN(e2.rowid)
  FROM events e2
  WHERE e2.event_id = e1.event_id
);
```

```
-- sessions 表去重
CREATE TABLE cleaned_sessions AS
SELECT s1.*
FROM sessions s1
WHERE s1.rowid = (
  SELECT MIN(s2.rowid)
```

```
FROM sessions s2
WHERE s2.session_id = s1.session_id
);
```

3.4 第 3 步：空值处理

```
-- 分类字段补默认值
UPDATE cleaned_users SET
  plan      = COALESCE(plan, 'free'),
  industry   = COALESCE(industry, '未知'),
  company_size = COALESCE(company_size, '1-10');

-- 数值字段补 0
UPDATE cleaned_subscriptions SET
  mrr = COALESCE(mrr, 0);
```

3.5 第 4 步：异常值过滤

```
-- duration_sec 超过 24 小时（86400 秒）视为脏数据
DELETE FROM cleaned_events
WHERE duration_sec > 86400 OR duration_sec < 0;

-- mrr 为负数视为异常
DELETE FROM cleaned_subscriptions
WHERE mrr < 0;

-- 漏斗标志位约束为 0/1
UPDATE cleaned_users SET
  is_activate = CASE WHEN is_activate NOT IN (0,1) THEN 0 ELSE is_activate END,
  is_pay       = CASE WHEN is_pay NOT IN (0,1) THEN 0 ELSE is_pay END,
  retain_7d    = CASE WHEN retain_7d NOT IN (0,1) THEN 0 ELSE retain_7d END;
```

3.6 第 5 步：时间格式标准化

```
-- 统一为 YYYY-MM-DD HH:MM:SS 格式
UPDATE cleaned_users SET
  signup_time = strptime('%Y-%m-%d %H:%M:%S', signup_time);

-- 过滤未来时间（注册时间不可能晚于当前时间）
DELETE FROM cleaned_users
WHERE date(signup_time) > date('now');
```

3.7 第 6 步：外键校验（删除孤儿记录）

```
-- events 表中的 user_id 必须存在于 cleaned_users
DELETE FROM cleaned_events
WHERE user_id NOT IN (SELECT user_id FROM cleaned_users);
```

```
-- sessions 表同理
DELETE FROM cleaned_sessions
WHERE user_id NOT IN (SELECT user_id FROM cleaned_users);

-- subscriptions 表同理
DELETE FROM cleaned_subscriptions
WHERE user_id NOT IN (SELECT user_id FROM cleaned_users);
```

3.8 清洗规则汇总

表	问题类型	清洗规则	处理方式
users	user_id 重复	主键去重	子查询保留最早记录
users	plan / industry 空值	缺失补默认值	COALESCE(plan, 'free')
users	漏斗标志位非 0/1	值域约束	CASE WHEN 归一
users	signup_time 格式不统一	时间标准化	strftime 格式化
users	未来时间注册	逻辑校验	DELETE
events	event_id 重复	主键去重	子查询保留最早记录
events	duration_sec 异常	值域过滤	> 86400 或 < 0 删除
events	user_id 不存在	外键校验	DELETE 孤儿记录
sessions	session_id 重复	主键去重	子查询保留最早记录
sessions	user_id 不存在	外键校验	DELETE 孤儿记录
subscriptions	mrr < 0	值域过滤	DELETE
subscriptions	mrr 空值	缺失补默认值	COALESCE(mrr, 0)

四、核心分析

4.1 五层转化漏斗建模

分析目的

构建「注册 → 激活 → 核心功能 → 付费 → 7 日留存」五层转化漏斗，定位每一步的转化率与流失率，找到最大卡点。

口径对齐原则

漏斗每一步的用户必须是上一步的子集，确保转化率计算的分子分母逻辑正确：

```
注册用户  激活用户  核心功能用户  付费用户  7 日留存用户
```

SQL 实现

```
WITH funnel AS (  
  SELECT  
    COUNT(*) AS total_users,  
    SUM(is_activate) AS activated,  
    SUM(CASE WHEN is_activate = 1 AND core_used = 1 THEN 1 ELSE 0 END) AS core_used,  
    SUM(CASE WHEN is_activate = 1 AND core_used = 1 AND is_pay = 1 THEN 1 ELSE 0 END) AS paid,  
    SUM(CASE WHEN is_activate = 1 AND core_used = 1 AND is_pay = 1 AND retain_7d = 1 THEN 1 ELSE 0 END) AS  
retained  
  FROM cleaned_users  
)  
SELECT  
  total_users AS '注册用户数',  
  activated AS '激活用户数',  
  core_used AS '核心功能用户数',  
  paid AS '付费用户数',  
  retained AS '7 日留存用户数',  
  ROUND(activated * 100.0 / total_users, 2) AS '激活率(%)',  
  ROUND(core_used * 100.0 / activated, 2) AS '激活→核心转化率(%)',  
  ROUND(paid * 100.0 / core_used, 2) AS '核心→付费转化率(%)',  
  ROUND(retained * 100.0 / paid, 2) AS '付费 7 日留存率(%)'  
FROM funnel;
```

漏斗分析结果

漏斗层级	用户数	阶段转化率	累计转化率
注册	5,000	—	100.00%
激活	3,160	63.20%	63.20%
核心功能	1,970	62.34%	39.40%
付费	660	33.50%	13.20%
7 日留存	269	40.70%	5.38%

关键发现

- **最大卡点：核心→付费转化率仅 33.50%。**每 3 个使用核心功能的用户中，只有 1 个转化为付费用户，说明核心功能的付费驱动力不足。
- **次大卡点：付费 7 日留存率 40.70%。**即便完成付费，仍有近 60% 用户在 7 天内流失，付费后的价值感知与持续使用存在断层。
- **激活率 63.20%** 有提升空间，近 37% 用户注册后未完成任何关键动作。

4.2 新手引导完成度与流失相关性分析

分析目的

验证假设：「新手引导完成度越高，用户流失率越低」。通过分箱对比量化引导完成度对留存的影响。

SQL 实现

```
SELECT
CASE
  WHEN guide_complete_rate = 0          THEN '0% 未开启'
  WHEN guide_complete_rate > 0 AND guide_complete_rate <= 0.3 THEN '1%-30%'
  WHEN guide_complete_rate > 0.3 AND guide_complete_rate <= 0.7 THEN '31%-70%'
  WHEN guide_complete_rate > 0.7 AND guide_complete_rate < 1  THEN '71%-99%'
  WHEN guide_complete_rate = 1          THEN '100% 全完成'
END AS guide_segment,
COUNT(*) AS users,
SUM(CASE WHEN retain_7d = 0 THEN 1 ELSE 0 END) AS churned,
ROUND(SUM(CASE WHEN retain_7d = 0 THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) AS churn_rate
FROM cleaned_users
GROUP BY guide_segment
ORDER BY churn_rate DESC;
```

分析结果

引导完成度	用户数	流失人数	流失率
0% 未开启	1,420	1,295	94.71%
1%-30%	680	459	67.50%
31%-70%	720	281	39.03%
71%-99%	660	171	25.91%
100% 全完成	1,520	733	48.26%

关键发现

- 引导未开启用户流失率高达 94.71%，几乎全部流失。
- 引导完成度 30% 是关键拐点：完成度从 0% 提升到 30%+ 时，流失率从 94.71% 骤降至 67.50%，激活率从 24% 跃升至 76%。
- 全完成引导的用户流失率仍为 48.26%，说明引导本身不能完全解决留存问题，还需配合核心功能体验与付费引导。

4.3 用户分层分析

分析目的

依据核心功能使用频次与付费状态，将用户划分为 S/A/B/C/D 五层，对比各层的付费率与留存差异，为精细化运营提供分层依据。

分层规则

层级	定义	运营含义
S	核心功能高频使用 + 已付费	高价值用户，需重点维护与扩展
A	核心功能中频使用 + 已付费	稳定付费用户，可引导升级
B	核心功能高频使用 + 未付费	高潜力用户，需推动付费转化
C	核心功能低频使用 + 未付费	流失风险用户，需激活干预
D	从未使用核心功能	冷启动用户，需引导触达

SQL 实现

```
SELECT
CASE
  WHEN core_used = 0 THEN 'D-冷启动'
  WHEN core_used = 1 AND is_pay = 1 AND freq >= 10 THEN 'S-高价值'
  WHEN core_used = 1 AND is_pay = 1 AND freq >= 3 THEN 'A-稳定付费'
  WHEN core_used = 1 AND is_pay = 0 AND freq >= 10 THEN 'B-高潜力'
  WHEN core_used = 1 AND is_pay = 0 AND freq >= 1 THEN 'C-待激活'
  ELSE 'D-冷启动'
END AS user_tier,
COUNT(*) AS users,
SUM(is_pay) AS paid_users,
SUM(retain_7d) AS retained_users,
ROUND(SUM(is_pay) * 100.0 / COUNT(*), 2) AS pay_rate,
ROUND(SUM(retain_7d) * 100.0 / COUNT(*), 2) AS retain_rate
FROM (
  SELECT u.*, IFNULL(e.freq, 0) AS freq
  FROM cleaned_users u
  LEFT JOIN (
    SELECT user_id, COUNT(*) AS freq
    FROM cleaned_events
    WHERE event_name = 'use_core_feature'
    GROUP BY user_id
  ) e ON u.user_id = e.user_id
) t
GROUP BY user_tier
```

ORDER BY pay_rate DESC;

分析结果

用户层级	用户数	占比	付费率	7 日留存率
S-高价值	180	3.60%	100.00%	78.30%
A-稳定付费	480	9.60%	100.00%	52.10%
B-高潜力	310	6.20%	0.00%	41.90%
C-待激活	1,040	20.80%	0.00%	18.30%
D-冷启动	2,990	59.80%	0.00%	5.40%

关键发现

- D 层（冷启动）用户占比高达 59.80%，近 6 成用户从未使用核心功能，是最大的流失池。
- S 层高价值用户留存率 78.30%，验证了「核心功能高频使用 + 付费」与留存的强正相关。
- B 层高潜力用户（高频使用但未付费）仅 310 人，是最值得投入付费转化资源的群体。
- C 层待激活用户留存率仅 18.30%，需通过运营触达提升核心功能使用频次。

4.4 渠道质量分析

分析目的

拆解各获客渠道的转化质量，为渠道预算分配与投放优化提供数据支撑。

SQL 实现

```
SELECT
  channel,
  COUNT(*) AS registrations,
  SUM(is_activate) AS activated,
  SUM(is_pay) AS paid,
  SUM(retain_7d) AS retained,
  ROUND(SUM(is_activate) * 100.0 / COUNT(*), 2) AS activate_rate,
  ROUND(SUM(is_pay) * 100.0 / COUNT(*), 2) AS pay_rate,
  ROUND(SUM(retain_7d) * 100.0 / SUM(is_pay), 2) AS paid_retain_rate
FROM cleaned_users
GROUP BY channel
ORDER BY pay_rate DESC;
```

分析结果

渠道	注册量	激活率	付费率	付费 7 日留存率
社群渠道	980	71.40%	18.37%	52.80%
内容营销	1,250	65.60%	14.20%	48.50%
搜索引擎	1,100	62.70%	12.80%	45.10%
信息流广告	1,670	58.40%	10.90%	38.20%

关键发现

- 社群渠道付费率 18.37% 最优，且付费 7 日留存率 52.80% 最高，是性价比最高的获客渠道。
- 信息流广告注册量最大（1,670）但付费率最低（10.90%），存在「量大质低」问题，需优化投放定向或缩减预算。
- 社群渠道应加大投入，信息流渠道应优化受众定向或降低出价。

五、问题总结

基于以上四维度分析，定位三大核心问题：

问题 1：新手引导触达不足

- 引导未开启用户占比 28.4%，其流失率高达 94.71%。
- 引导完成度 30% 是留存拐点，但大量用户未跨过此门槛。
- 根因：引导入口不明显、引导内容过长导致中途放弃、未与用户核心诉求关联。

问题 2：核心价值感知滞后

- 59.80% 用户从未使用核心功能（D 层冷启动）。
- 核心→付费转化率仅 33.50%，说明核心功能的付费驱动力不足。
- 根因：核心功能入口隐藏、首次体验缺乏 Demo 引导、付费价值未在前 3 天传达。

问题 3：流量未精细化运营

- 信息流渠道注册量最大但付费率最低（10.90%），预算浪费严重。
- 社群渠道质量最优但投入不足。
- 各渠道采用统一运营策略，未按渠道特征差异化触达。

六、落地策略

6.1 产品策略

策略	优先级	预期效果	对应问题
新手引导改版：Step 指示器 + 价值前置 + AI Demo 卡片	P0	中途跳出率 -22%，CTA 点击率 +18%	问题 1
核心功能入口前置：首屏可见 + Demo 页一键体验	P0	核心功能采用率 +15%	问题 2
导航架构改版：顶部双层 Tab + 60px 图标轨 + 全局搜索	P1	功能可发现性 +22%，误用率 -24%	问题 2
激励弹窗系统：升级 Pro + 签到 + 成就 + 邀请 4 场景	P1	付费转化 +48%，7 日留存 +35%	问题 1/2

6.2 运营策略

策略	优先级	预期效果	对应问题
分层运营：S/A 层 1v1 服务，B 层付费 Trial，C 层激活触达，D 层再唤醒	P0	D 层转化 5%+，B 层付费率 +20%	问题 2
渠道预算重分配：社群 +30%，信息流 -20%，释放预算至内容营销	P0	整体付费率 +2pp	问题 3
签到机制：连续签到积分 + 第 7/14/30 天梯度奖励	P1	7 日留存 +35%	问题 1
好友邀请：双方各得 ¥60 Pro 会员金	P2	N 因子 2.4x，获客成本 -46%	问题 3

6.3 Figma 原型改版设计稿

基于上述策略，输出 4 张高保真 Figma 原型改版设计稿：

设计稿	场景	核心改版策略
新手引导首屏	注册后首次进入	Step 指示器 + 价值前置 + AI 助手实时 Demo
核心功能 Demo 页	首次体验 workflows 编辑器	三栏式编辑器 + 多模型智能路由 + 节点状态可视化
导航对比 V1→V2	全局导航架构改版	顶部双层 Tab + 60px 图标轨 + K 全局搜索
激励弹窗系统	4 场景全覆盖	升级 Pro + 每日签到 + 成就解锁 + 好友邀请

七、项目成果

7.1 产出物清单

产出物	数量	说明
Notebook 代码	8 个 Cell	从数据导入到分析导出的完整流程
原始数据集	5 张 CSV	users / events / sessions / subscriptions / saas_user_funnel
清洗后数据集	5 张 cleaned 表	cleaned_users / cleaned_events / cleaned_sessions / cleaned_subscriptions / cleaned_funnel
分析结果数据	9 份 CSV	漏斗 / 引导相关性 / 用户分层 / 渠道质量 等
业务复盘报告	1 份	本文档
Figma 原型设计稿	4 张	新手引导 / 核心功能 Demo / 导航对比 / 激励弹窗
可分享网页版	1 份	单文件 HTML，零外部依赖

7.2 能力沉淀

本项目沉淀了从数据洞察到产品运营落地的完整链路能力：

数据生成 → SQL 清洗 → 漏斗建模 → 用户分层 → 渠道分析 → 问题定位 → 策略输出 → 原型设计 → 网页部署

每一步均产出可复用的方法论与代码资产，可直接迁移至真实业务场景。

八、附录

附录 A：关键指标速查

指标	数值	说明
总用户数	5,000	模拟生成
激活率	63.20%	注册→激活
核心功能采用率	39.40%	注册→核心功能
付费率	13.20%	注册→付费
7 日留存率	5.38%	注册→7 日留存（累计）
核心→付费转化率	33.50%	漏斗最大卡点

付费 7 日留存率	40.70%	漏斗次大卡点
引导未开启流失率	94.71%	最强流失因子
引导全完成流失率	48.26%	仍有改善空间
引导完成度拐点	30%	流失率骤降分界线
冷启动用户占比	59.80%	从未使用核心功能
社群渠道付费率	18.37%	渠道最优
信息流渠道付费率	10.90%	渠道最差

附录 B：来源声明

本项目来源于和鲸社区，使用转载需要标注来源。

- 作者：角岛鲸 nq2v
- 来源：<https://www.heywhale.com/mw/project/6a8169684e551f2d69b5609f?shareby=6a81643b24ec58806f3f27df#>
- 平台：和鲸社区（ModelWhale）
- 数据性质：模拟数据（Python Faker + NumPy 生成）
- 输出日期：2026 年 08 月 16 日

报告完