

PPC广告投放ROAS优化项目报告

一 项目摘要

项目背景

公司正在进行大规模的 PPC 广告投放（约 1000 条活动）。过去一年多的投放数据显示：

- **投放回报不稳定：** ROAS 呈现高波动，导致预算投入无法提前预估产出风险
- **决策缺失：** 高 ROAS 活动特征未沉淀，经验不可复用。缺乏量化标准，优质活动放量不足，低效活动仍消耗
- **预算使用效率低：** 80% 活动 $\text{ROAS} < 18$ ，大量资金流向低效投放。

项目目的

通过分析广告活动的表现，识别高 ROAS 路径，来提升后续的 ROAS 表现

- 识别影响 ROAS 的关键指标
- 挖掘高 ROAS 活动的稳定结构化路径
- 量化当前投放效率与潜在优化空间
- 输出可执行的策略建议及对应的 A/B 测试方案，验证因果效应

核心发现

- CVR、RPC、CPC 为影响 ROAS 的核心变量
- CPC 0–5 为关键区间
- 基于决策树挖掘出两条可规模化的高 ROAS 路径

最终策略建议

优先加大对 CPC 0–5 区间的投放占比

聚焦 **路径一与路径二** 的投放组合进行扩量优化

预期价值提升

优化后整体 ROAS 预计可提升约 **30%–60%**（来自价值量化部分）

二 数据概览

本次分析的数据集共包含 1000 条广告活动在各类数字平台上的记录，涉及 19 个字段，覆盖活动标识、广告特征、投放地域、投放结果与关键绩效指标。

数据来源：<https://www.kaggle.com/datasets/aashwinkumar/ppc-campaign-performance-data>

(一) 字段介绍

	字段	描述
主键	Campaign_ID	每个广告活动的唯一标识符。
类别变量	Platform	活动投放的平台（如 Instagram、LinkedIn、YouTube、Facebook、Google）。
	Content_Type	广告内容形式（Image / Video / Carousel / Text）。
	Target_Age	目标受众年龄段（如 18-24、25-34）。
	Target_Gender	目标受众性别（Male / Female / Other）。
	Region	活动覆盖的地理区域（North America / Europe / Asia 等）。
数值变量	Impressions	广告曝光次数。
	Clicks	用户点击次数。
	Conversions	完成目标行为的数量（如注册、购买）。
	CTR (Click-Through Rate)	点击率 = 点击数 / 曝光数。
	CVR (Conversion Rate)	转化率 = 转化数 / 点击数。
	CPC (Cost Per Click)	每次点击成本 = Spend / Clicks。
	CPA (Cost Per Acquisition)	每次转化成本 = Spend / Conversions。
	Budget	为该活动分配的预算（以美元计）。
	Spend	实际总花费。
	Revenue	转化带来的总收入。
	ROAS (Return on Ad Spend)	投资回报 = Revenue / Spend，代表广告支出产生收入的效果。
	Duration	广告系列持续在线的天数
时间变量	Date	数据记录日期，用于观察活动随时间的行为或趋势。

新增字段：

- **RPC (Revenue Per Conversion)**：单次转化带来的平均收入 = Revenue / Conversions。
 - 注意：本数据中的“转化”包含注册与购买等不同类型的行为，因此 RPC 不等同于传统意义上的客单价 (AOV)。

(二) 数据清洗与预处理

本项目的原始数据存在字段类型不一致、计算逻辑异常、业务规则异常等问题。

清洗步骤如下：

- **字段规范化**
- **指标重新计算**
- **业务规则异常剔除:** 剔除 26 条业务逻辑异常 (如: Clicks > Impressions, CTR > 100%)
- **异常值与log转化:** ROAS、Revenue 等核心指标呈现显著的右偏分布。采用 log1p 转换

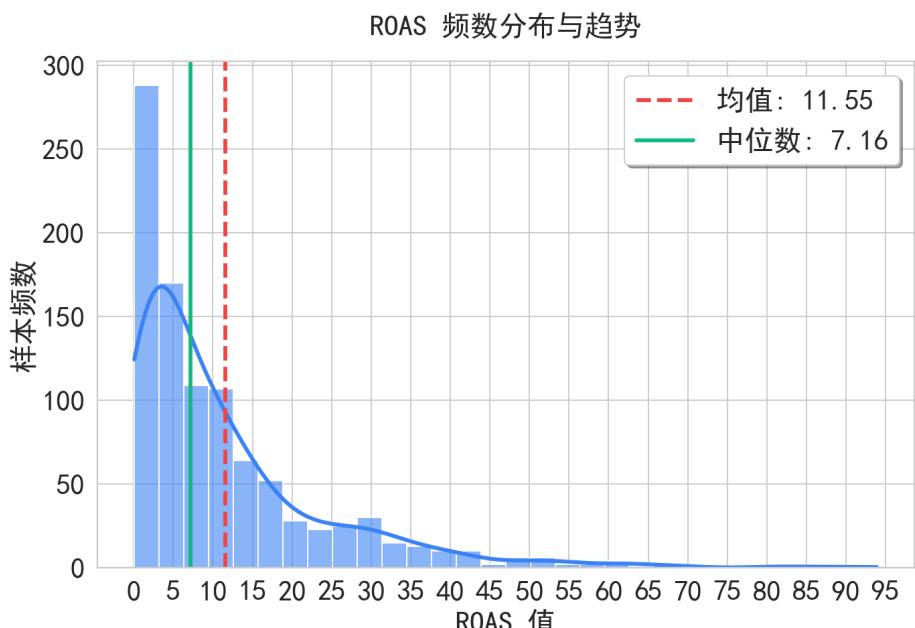
总结: 清洗后的数据共 974 条, 满足后续分析与模型构建要求。

三 核心发现

背景: ROAS 呈显著右偏分布

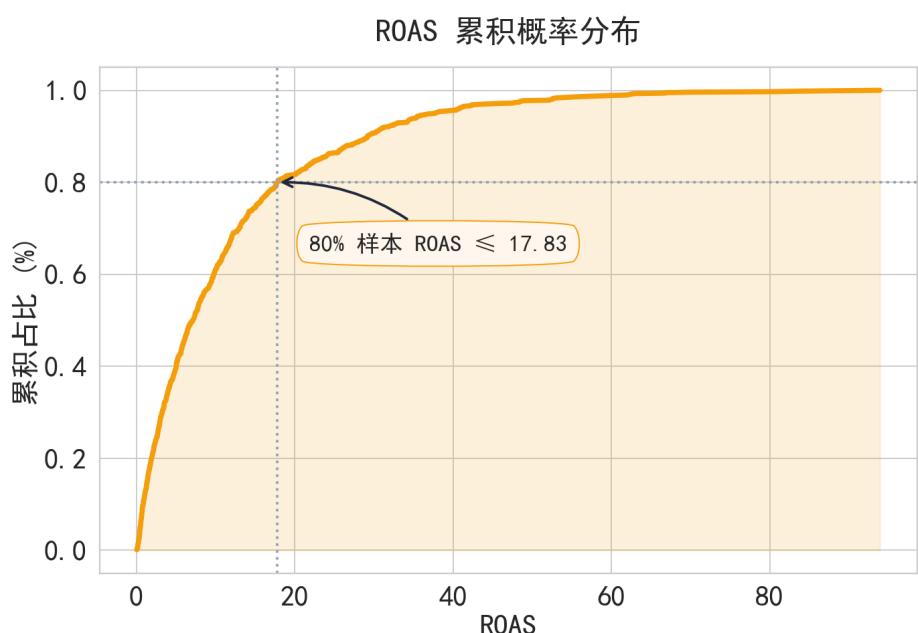
关键统计指标

- 均值: 11.55
- 中位数: 7.16
- 标准差: 13.1



分布特点

- ROAS 呈显著**右偏**, 存在长尾高值
- 约 80% 的活动 **ROAS < 18**, 说明大部分预算集中在低效活动
- 典型**帕累托现象**: 少数高效活动贡献主要回报



(一) 分类变量对于ROAS影响有限

通过对平台、内容类型、年龄、性别和地区的样本分布、 ROAS 中位数分析、 ROAS 标准差和中位数四象限图分析发现：

- 不同类别间 ROAS 差异较小 (ROAS中位数差异仅在 1-3 点左右)
- 不同分类变量的标准差和ROAS中位数四象限分布也未显示显著趋势

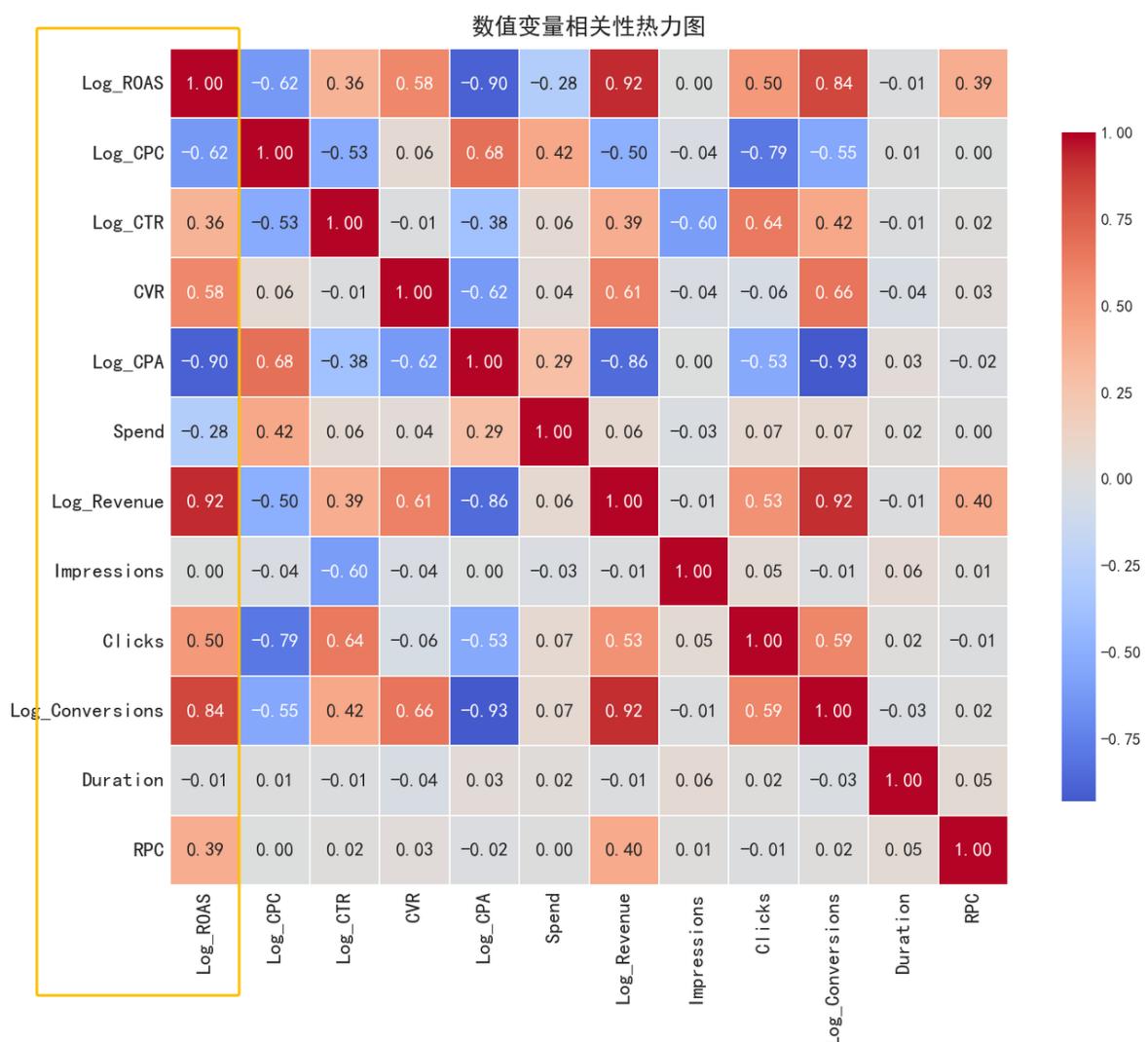
结论：分类变量在整体收益表现中贡献有限，可不作为优化重点。

后续随机森林模型的特征重要性也验证了这一点：所有类别变量的重要性均低于 0.01，说明其对 ROAS 的解释能力非常有限。

(二) CPC、CVR、RPC是roas的核心驱动指标

1. ROAS的关键数值因素：CPC、CVR、RPC

Pearson 相关分析显示：CVR (0.58) 、 RPC (0.39) 、 CPC (-0.62) 与 ROAS 关联紧密。



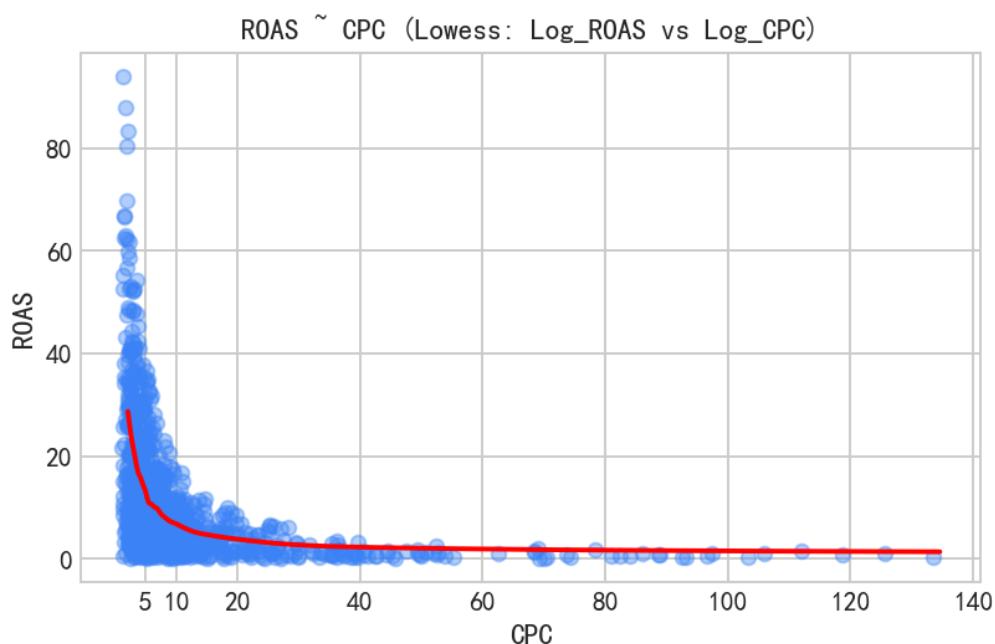
- CPA 与 ROAS 相关性更高 (-0.9) ，但 CPA 难以直接调控，故重点关注 CPC。

- CTR 在全量数据相关性为 0.36，但在 $CPC \leq 5$ 区间内与 ROAS 无明显关系，说明 CTR 对 ROAS 的影响主要通过 CPC 间接体现，后续分析不再关注。

2.CPC关键成本杠杆，且CPC0-5为关键高效区域。

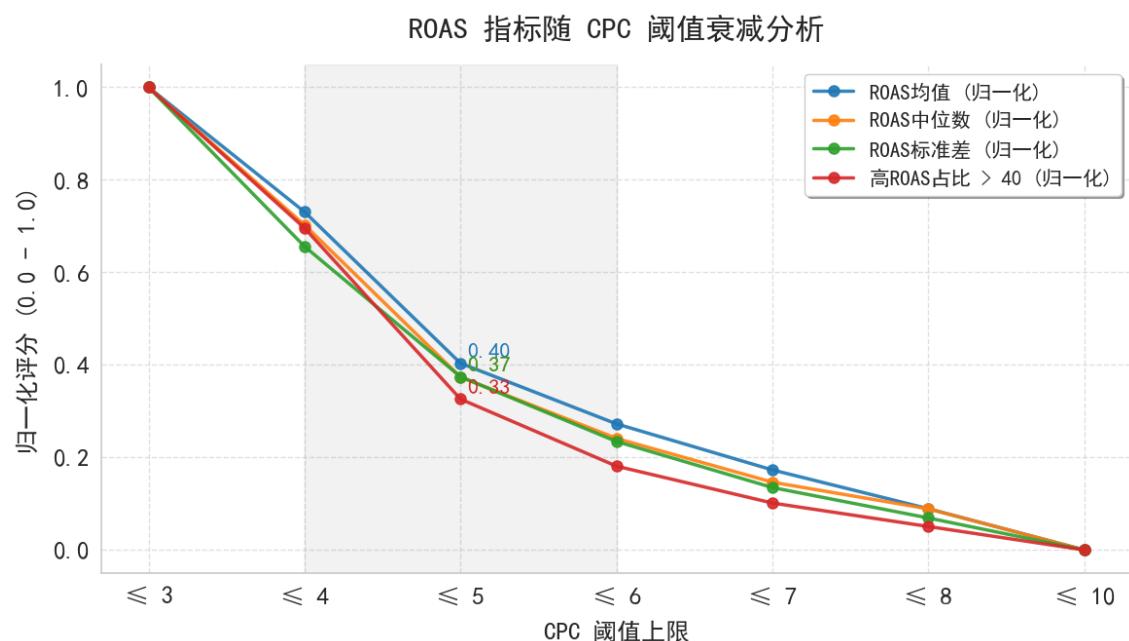
2.1 CPC关键成本杠杆

ROAS 随 CPC 增加呈指数下降，低 CPC 区域效率最高。



2.2 CPC(0-5]为高效关键区间

为锁定最优出价区间，下文结合 ROAS 指标衰减趋势（图）与 CPC 分位区间统计（表）进行对比分析：



CPC 范围	ROAS 均值	ROAS 中位数	样本量	样本占比	业务定义
3.26 - 4.20	16.75	13.89	97	约 10%	维持极高回报
4.20 - 4.94	12.5	11.1	98	约 10%	收益最后堡垒
4.94 - 5.89	12.08	9.98	97	约 10%	中位数跌破10.0
5.89 - 7.26	9.71	8.33	97	约 10%	正式进入低效区
CPC ≤ 5	-	-	488	41.20%	占据业务半壁江山

高效区间（核心逻辑）

- ROAS 随 CPC 增加呈指数型下降，低 CPC 区域天然具备最高投放效率。

黄金拐点（高敏区）

- CPC ≤ 5 是收益衰减由陡转平的拐点。
- 在 4.20–4.94 的区间，ROAS 中位数稳定在 11.1；超过 5 的区间，中位数首次跌破 10.

核心利润来源（高质流量）

- ROAS > 40 的样本仅存在于 CPC ≤ 5 的低成本区间，且占比为高成本区的 1.7 倍。
- 这些高回报活动是投放利润的核心，“黄金区间”须保留，防止后续低效流量稀释整体业绩。

数据代表性强（基本盘）

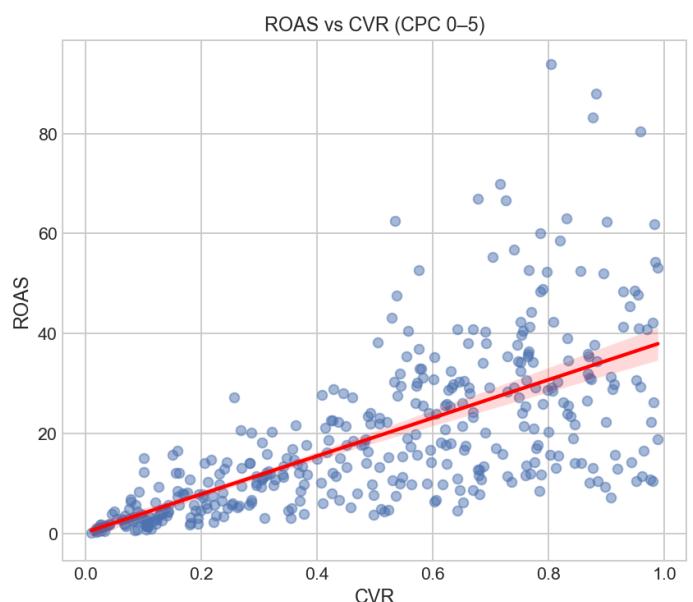
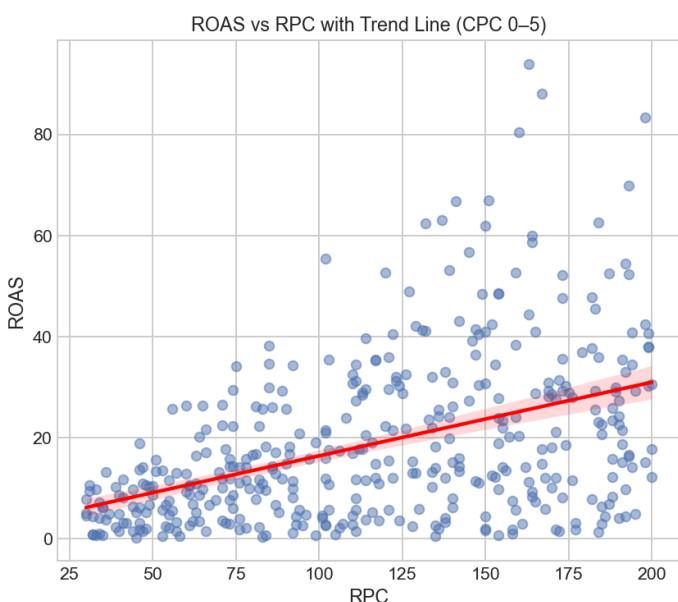
- CPC ≤ 5 的样本量占比 41.2%，并且贡献了 55.3% 收入。

3. CVR RPC 对 ROAS 起到关键推动作用，并且两者为协同作用

3.1 CVR 与 RPC 对 ROAS 呈显著正向影响

为了控制成本影响，同时聚焦高效投放区间，本分析限定 CPC ≤ 5。

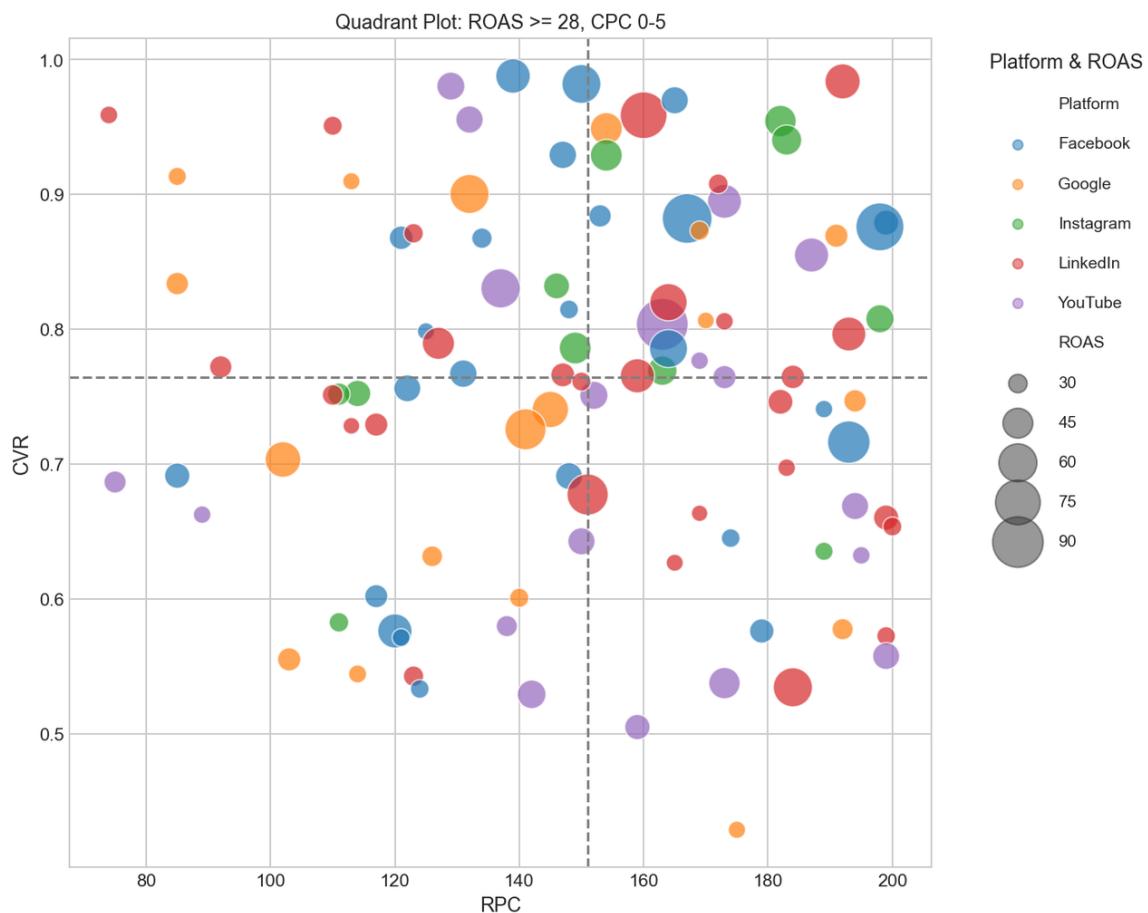
CVR 与 RPC 对 ROAS 呈显著正向关系：



3.2 CVR与RPC为协同作用推动ROAS

通过四象限分析低 CPC (0-5) 区域高 ROAS ($\geq 75\%$ 分位数) 的活动特征发现：

- 数据点主要集中在右上象限，表明高 ROAS 活动通常同时具有较高 CVR 和 RPC
- 说明 CVR 与 RPC 协同推动 ROAS



(三) 基于决策树的高 ROAS 策略路径挖掘

为将复杂的数据转化为可执行的投放指令，本项目通过决策树算法，针对不同经营目标挖掘核心规则。

1. ROAS区间划分：选取10-20与20-40，聚焦高价值广告活动。

为了兼顾 策略稳定性和高收益机会洞察，将 ROAS 划分为两个业务区间：

- **稳态优化区（10-20）**：样本量大（202条）、波动小（标准差约1.3），用于提取可复制的稳健投放规则。
- **高收益探索区（20-40）**：样本量充足（136条）、ROAS中位数28，多为低CPC+高CVR+高RPC组合，用于识别潜在高回报策略。

2. 核心路径识别与回测

经过算法挖掘到六条路径，后全量数据回测（剔除样本量极小以及ROAS表现不好的失效路径）

锁定两条核心策略：

	路径	核心规则定义	样本量	ROAS均值	ROAS标准差	决策	理由
第一组 ROAS (10- 20]	Path 1	RPC ≤ 180.50 且 CVR > 0.48 且 Target_Age 非 25-34	350	14.97	13.57	√	虽然样本量大，但ROAS均值不高且相对波动较大。
	Path 2	RPC > 180.50 且 CPC ≤ 9.85	98	22.78	15.51	√	高收益路径，样本量中等，可小规模投放探索
	Path 3	RPC > 180.50 且 CPC > 9.85 且 Target_Age 为 35-44	4	5.51	4.33	✗	样本量太少
第二组 ROAS (20- 40]	Path 4	CVR > 0.53 且 CPC ≤ 5.29 且 RPC > 109.00	117	37.77	15.32	√	ROAS表现最顶尖，属于高转化、低成本的“极致效率”模型。优先投放，重点策略
	Path 5	CVR > 0.53 且 CPC > 5.29 且 CTR ≤ 0.21	209	8.16	6.9	✗	ROAS均值太低
	Path 6	CVR ≤ 0.53 且 CTR > 0.68 且 RPC > 158.00	7	11.98	9.29	✗	样本量太少

核心路径

- **路径一（稳态区）**：RPC > 180.50 且 CPC ≤ 9.85
 - 业务含义：确保在高价值转化前提下，控制点击成本，适合大盘常态化投放。
- **路径二（高收益区）**：CVR > 0.53 且 CPC ≤ 5.29 且 RPC > 109
 - 业务含义：极致压缩点击成本并筛选高转化率素材，适合作为爆款活动扩量模型。

3. 路径验证：随机森林

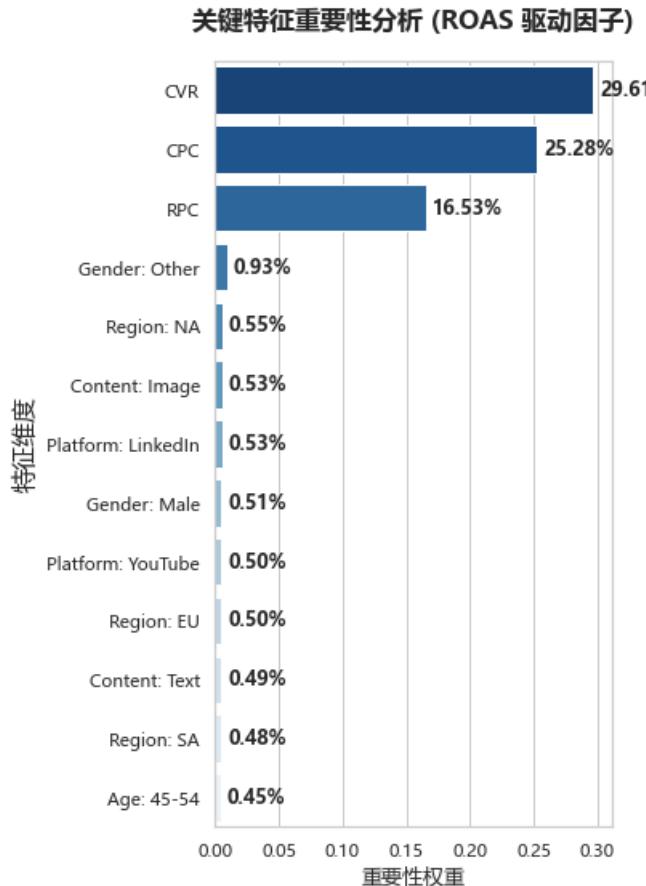
为克服决策树易过拟合的局限性，本分析引入随机森林进行特征重要性交叉验证：

特征重要性显示：

- CVR、CPC、RPC为主要驱动因素
- 分类变量（平台/内容类型/受众）重要性 <0.01 ，说明对ROAS影响有限

结论：

交叉验证确认了数值变量是驱动ROAS的关键，证明了决策树路径在业务指导上的科学性与稳健性。



(四) 价值量化 - 当前状态 vs 优化潜力

目标：量化当前广告活动价值，并评估通过优化策略提升 ROAS 的潜力。

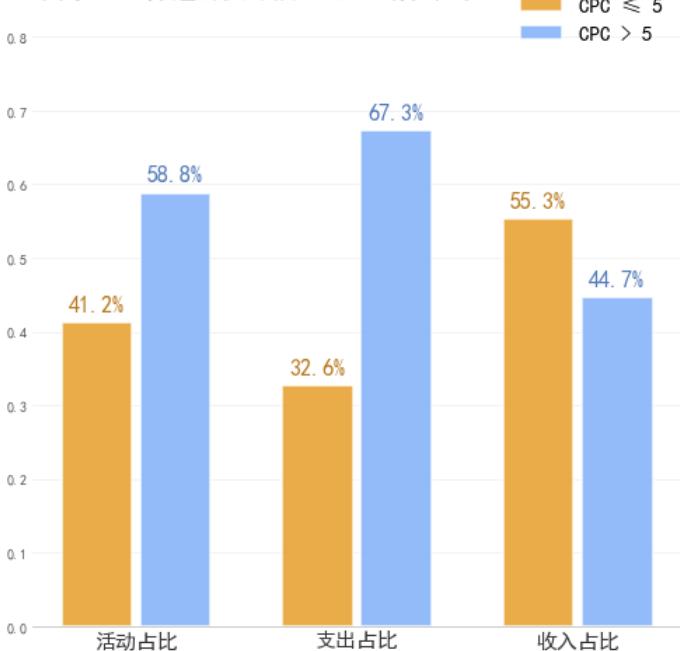
1. 当前状态分析

整体情况：

- 活动总数：974
- 总支出：5,792,405
- 总收入：57,560,796
- 平均 ROAS：11.55，中位数 7.16

2. CPC 区间分布 (≤ 5 vs > 5) :

不同 CPC 分组的资源投入与产出分布对比

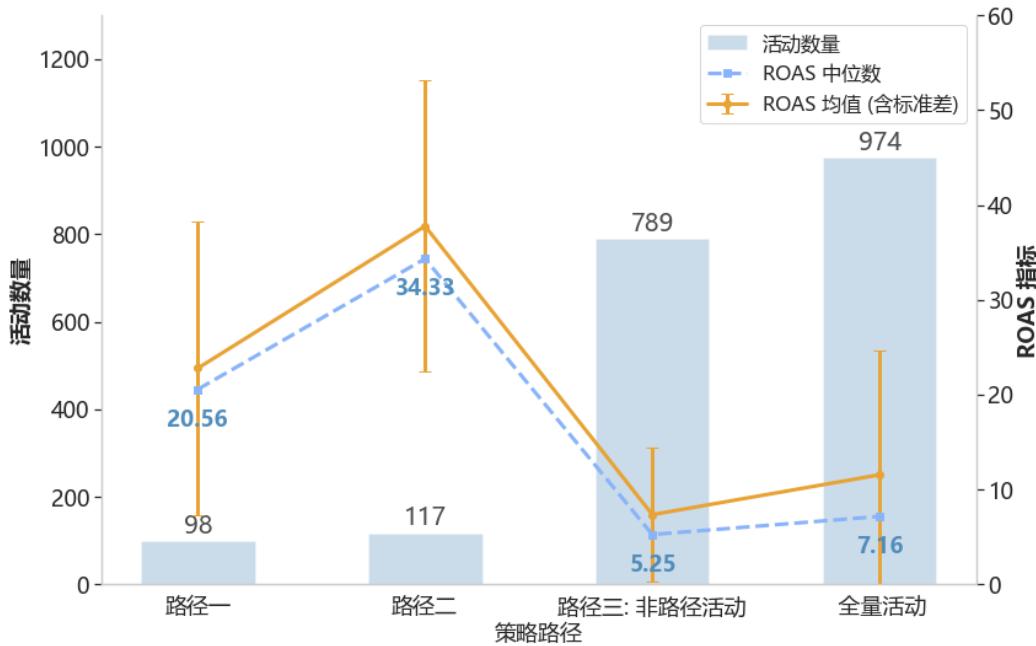


- CPC ≤ 5: 占比41%，贡献总收入55%
- CPC > 5: 占比59%，贡献总收入45%

低成本活动虽然占比少，但贡献收入更高，值得关注。

3. 核心路径表现：

各路径活动规模与 ROAS 表现对比



洞察：

路径一和路径二的 ROAS 明显高于非路径活动，是优化重点。

4. 优化场景模拟

场景 1：停止低效活动，聚焦 CPC ≤ 5

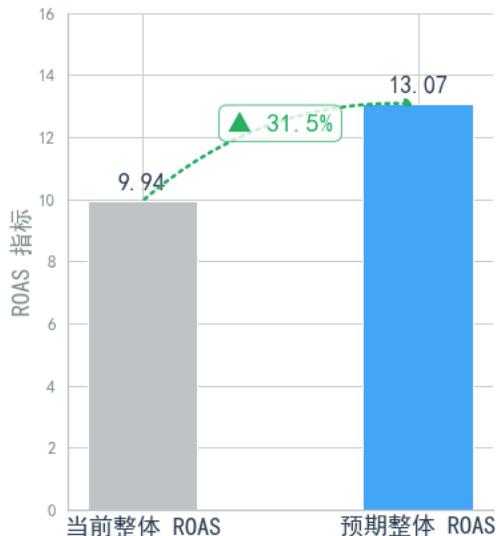
- 可重新分配预算：3,901,577
- 假设新增 ROAS:

中位数 \times 80% \rightarrow 11.25

- 预期整体 ROAS: 13.07 (提升 31.5%)

聚焦低成本活动能显著提升整体效率

场景 1: 预算重分配后的 ROAS 提升预测



优化策略: 停止 CPC>5 活动

- 释放预算: \$3,901,577
- 预期增收: \$43,884,937

场景 2: 复制两条核心路径 (50%预算投放)

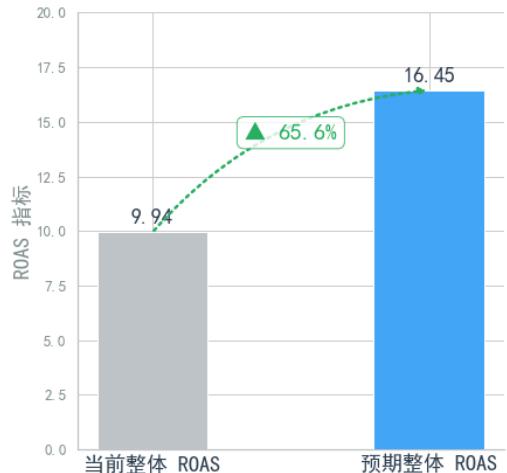
- 路径活动 ROAS:

中位数 \times 80% \rightarrow 22.97

- 预期整体 ROAS: 16.45 (提升 65.6%)
- 预期收入: 95,300,372 (增加 37,739,576)

聚焦高 ROAS 路径能最大化投放收益

场景 2: 优质路径结构性加码后的 ROAS 提升预测



优化策略: 结构性加码优质路径

- 加码预算: 50% (基于优质路径)
- 预期增收: \$37,739,576

5. 小结

- 当前状态: 低CPC活动贡献高收入, 核心路径ROAS表现突出
- 优化场景模拟表明:
 - 剔除低效活动 \rightarrow ROAS提升约30%
 - 聚焦两条核心路径 \rightarrow ROAS提升约65%
- 策略洞察:
 - 日常优化·聚焦低成本高转化活动
 - 核心高 ROAS 路径可作为重点策略组合进行预算倾斜

(五) 策略验证：A/B 测试设计

目标：验证决策树模型识别出的两条高 ROAS 路径在实际投放中是否具备可复现的因果效应，为预算分配提供依据。

1. 实验背景

- 分析约 1000 条 PPC 广告活动发现：
 - ROAS 呈帕累托分布（均值 11.55，中位数 7.16），少数高 ROAS 活动贡献主要收益
 - 高 ROAS 活动存在明确数值特征（CPC、CVR、RPC 协同作用）
- 决策树模型识别出两条稳定高 ROAS 路径，但仅为相关性，需要 A/B 测试验证因果效应。

2. 核心业务假设

路径	模型条件	业务逻辑	落地策略
路径一：稳态高收益组合	RPC > 180.50 且 CPC ≤ 9.85	高 RPC 表示该组合平均收益高 低 CPC 保证成本可控	- 设置出价上限 (Bid Cap) ≤ 9.85 - 优先分配预算给历史高收益活动组合 - 监控 ROAS 与单次转化收益，调整预算权重
路径二：高转化效率组合	CVR > 0.53 且 CPC ≤ 5.29 且 RPC > 109	高 CVR 表示内容/受众匹配度高， 低 CPC 控制成本	- 优先投放历史高转化受众或相似受众 - 使用表现良好的素材（如视频） - 设置合理出价上限，控制 CPC 在可控范围 - 聚焦提升转化率和转化价值，而非单纯追求低 CPC

3. 实验设计

- 组别：**
 - Control：现状（自动出价，素材混合）
 - Test A：路径一策略，Bid Cap ≤ 9.85，素材保留历史 RPC 高于平均的 Top 3
 - Test B：路径二策略，Cost Cap ≤ 5.29，素材针对高 CVR 历史受众。
- 随机化：**User ID 哈希分流，保证同一用户仅在一个组
- 流量分配：**Control 34% / Test A 33% / Test B 33%
- 实验对象：**过去 30 天活跃可转化受众

4. 指标体系

- 核心指标 (Primary Metric) :** ROAS
 - Test A/B 必须显著高于 Control（提升幅度 > 15%）
- 次级指标 (Secondary Metrics) :** CVR、RPC、CPC

- 护栏指标 (Guardrail Metrics) : Spend \geq 对照组 80%，监控 CPM，避免极端压价
-

5. 样本量 & 统计方法

最小样本量：

- 设显著性水平 $\alpha=0.05$ ，功效 Power=0.8
- 为检测 15% ROAS 提升，每组需累积约 1,000 次转化 (Conversion)

流量推演：

- Test A (路径一组) : 需约 11,411 次曝光
- Test B (路径二组) : 需约 9,146 次曝光

实验周期与监控：

- 当各组 Conversion $\geq 1,000$ 次后，进行 P 值检验
- 若 $P < 0.05$ 且 ROAS 提升达标，则判定路径因果效应成立，可转入全量上线决策

统计方法：

- ROAS 偏态处理：99 分位 Winsorization
- 不确定性估计：Bootstrap 重采样
- 组间显著性检验：Mann-Whitney U (非参数)

样本不足应对策略：

- 延长实验周期
 - 扩大受众覆盖
 - 使用 CVR、CPC 中短期趋势作为早期指示
-

6. 停机准则 & 决策矩阵

停机准则：

条件	动作
ROAS 连续 3 天低于对照组 30%	立即关停
流量异常（无消耗）或技术故障	暂停/恢复为基准

实验决策矩阵：

实验结果	决策
ROAS 显著提升且消耗稳定	全面上线
ROAS 提升但消耗下降明显	保留为小众策略
无显著差异	迭代假设，重新建模
ROAS 下滑或消耗失败	废弃策略

四 总结

(一) 关键洞察

1. 高 ROAS 活动特征

- CVR、RPC、CPC是驱动 ROAS 的主要因素，分类变量（平台、内容类型、年龄、性别、地区）对 ROAS 的区分力有限。
- 低 CPC 区间（0–5）为关键区域。覆盖大部分样本，高 ROAS 活动集中于此，回报随 CPC 升高呈现收益衰减趋势。

2. 高ROAS路径：两条核心策略路径：

- **路径一（稳态区）**：CPC ≤ 9.85 且 RPC > 180.50
- **路径二（高收益区）**：CPC ≤ 5.29 且 CVR > 0.53 且 RPC > 109

(二) 最终策略

1. 投放优化方向

- **重点关注**提升 CVR 和 RPC，同时保持 CPC 在合理区间。
- **稳态优化**：采用路径一策略，保证基础收益稳定。
- **高收益探索**：采用路径二策略，寻找潜力高回报活动。

2. 资源分配

- 低 CPC 区间（0–5）优先投放，高 ROAS 活动密集，可作为核心投放区。
- 控制预算，保证实验组/策略组消耗 \geq 对照组 80%，监控 CPM 避免极端压价。

(三) 对业务建议

- 使用低 CPC + 高 CVR + 适中 RPC 的组合进行日常投放优化。路径规则可直接作为预算分配、素材选择和目标受众筛选的参考依据。
- 分类变量可作为辅助参考，但不应作为主要决策因素。
- 对核心路径进行 A/B 测试验证因果效果，监控核心指标（ROAS、CVR、RPC）及护栏指标（CPM、Spend），避免数据偏差。

(四) 下一步工作

1. 持续优化

- 随业务变化动态调整路径条件，持续监控 ROAS、CPC、CVR 和 RPC。
- 对新素材、新受众进行策略验证，保持路径策略有效性。
- 补充素材标签等数据，量化创意对CVR影响。

2. 探索更多优化手段

- 尝试使用机器学习预测模型进行高 ROAS 活动预测，辅助决策树路径。
- 分析其他潜在影响因素（如投放时间段、创意形式）对 ROAS 的作用。

3. 业务闭环

- 将策略执行效果反馈到分析模型，形成数据驱动的投放优化闭环。

五 不足与局限

数据维度有限：缺少细致内容信息、用户画像和上下文指标，无法深入分析创意或受众对ROAS的具体影响。核心数值指标（CVR、RPC、CPC）虽然可以识别关键驱动因素，但无法拆解背后的具体原因。

路径可执行性受限：决策树条件不可完全操控，样本量有限，稳健性不足。

分析深度有限：优化路径数量有限，数据更多用于识别趋势而非完全可操作策略。

实验假设依赖性：A/B测试中对ROAS扩量等假设可能与实际业务存在偏差，增加策略落地不确定性。