

数据期 2023 年货节 / 情人节

34,893 销售 · 4,339 退款 · 37,469 咨询

SQL + Python + jieba NLP

01 背景

- 唇釉大促用「第 2 件半价、第 3 件低至 20 元」吸引凑单，但结算底层实为「总价打折、均价分摊」，宣传价与结算价对不上。
- 活动期 GMV 大涨，退款率与客诉却同步飙升——怀疑「GMV 虚假繁荣」，需要量化剪刀差并定位根因。

02 数据

- 三份原始数据：销售 34,893 行、退款 4,339 行、客服咨询 37,469 行，覆盖年货节与情人节两期大促。
- 口径：订单级退款率 = 退款订单 ÷ 全部订单；实收 GMV = GMV - 退款金额；规则类退款 = 「价格/优惠规则看不懂」类。
- 方法：SQLite 装载派生 + 16 组 SQL 聚合 + Python (3 σ 异常检测、GMV 增量归因、卡方/z 检验、jieba 情感挖掘)。

03 发现

- 「虚假繁荣」成立：GMV +56.8%，实收仅 +51.6%（剪刀差 5.2pct）；订单级退款率 8.48% → 15.42%（ $z=19.4$, $p<0.001$ ）。
- 根因是规则复杂度而非产品质量：规则类退款占比 43.24% → 83.05%，质量类原因不升反降。
- 客诉压力放大：愤怒词 ×2.8、困惑词 ×4.1、强负面咨询占比 5.81% → 20.19%。
- 机制闭环：凑单诱导（3 件装 3.7%→31%）→ 均价分摊实付 80 元（实际折扣 27.3%，非宣传 80%+）→ 感知欺骗 → 退款客诉 → 实收承压。

04 建议

- 短期止血：上线「买 X 件立省 Y 元」计算器 + 客服安抚话术，先压客诉与退款。
- 中期重构：3.8 大促改为「均价可见」表达，让宣传与结算口径一致。
- 长期预警：退款原因—咨询类别双通道 3 σ 预警，提前发现大促数据异常。

05 结果

- 24 小时内定位异常并量化「GMV 虚假繁荣 vs 实收惨淡」的剪刀差，把隐性问题变成可汇报结论，避免营销预算持续投在错误渠道。
- 边界：两期分属不同节日档，只能证「显著相关」、不能证因果唯一；下一步用同期 AB 实验升级为因果证据。

业务价值

用异常检测提前发现大促数据波动，24 小时定位问题，避免多花 6 位数营销预算在错误渠道。