

单表数据先验 · 脱敏交付样例

先看清问题，再决定怎么改。

这不是“把 Excel 排版好看”的样例。它展示一次 199 元先验会怎样把错漏、业务影响、修复顺序和不能自动处理的部分写成可执行判断。

合成数据

场景演绎，不对应真实客户或项目

单表上限

5,000

行

字段上限

30

列

书面判断

48h

内交付

SYNTHETIC ORDER TABLE

462

个问题项

11.6%

异常行占比

订单号	日期	金额	
DEMO-1001	09/28	1280	
DEMO-1001	09/28	1280	●
DEMO-1003	13/02	640	●
DEMO-1005	10/01	1,200元	●
DEMO-1013	10/09	1.2e3	●

P0 日期不存在 / 倒置

P0 重复订单号

P1 金额混入文字

P1 客户号空值

P2 门店名称不统一

SAMPLE REPORT / NOT CUSTOMER DATA

先建立一张能复核的表况底稿

先验不急着给“清洗方案”。先把表是什么、问题有多少、哪些字段影响汇总和交接说清楚。

● 总行数

4,280

合成订单表

● 总列数

24

识别 11 个关键字段

● 问题项

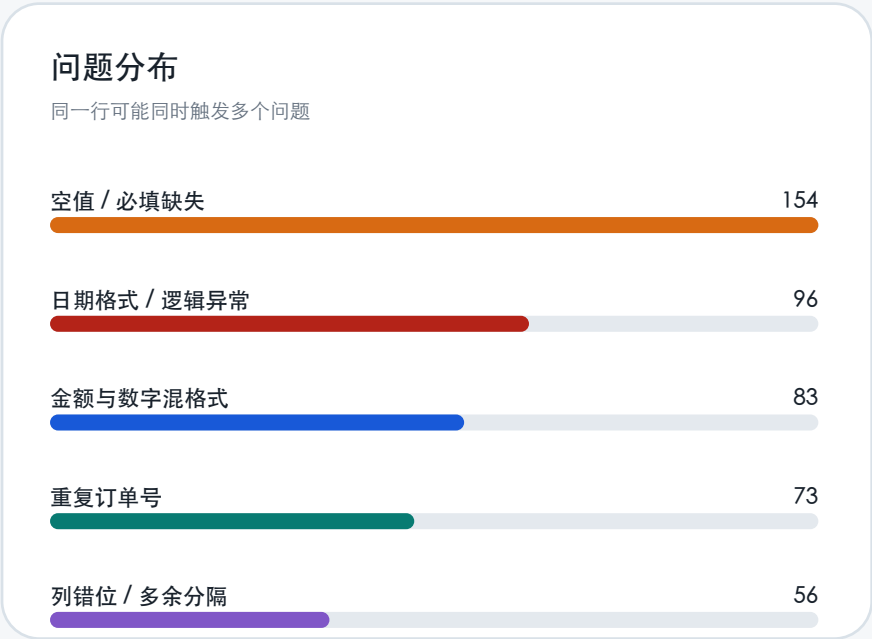
462

按问题项计数

● 异常行

497

约 11.6%，可重叠



一句话判断

这张表不是不能用，而是不能直接拿去做月度汇总。

日期、金额和订单主键会直接影响合计与去重结果。先确认业务规则，再决定自动修、人工补，还是退回来源系统。

不能直接自动覆盖原值

本样例的输入结构

字段	类型	先验关注点	示例
订单号	文本	主键重复、空值、前后空格	DEMO-1001
下单日期	日期	不存在、倒置、日月顺序	2026/13/02
金额	数值	货币符号、千分位、科学计数法	1.2e3
客户号	文本	必填缺失、跨表关联完整性	C-031 / 空
门店	分类	同义值、前后空格、区域映射	东区 / 东门

每个问题都要能指回字段和影响

“看起来有点乱” 不能作为交付。先验会把问题分成可定位、可解释、可确认的条目。

优先问题清单

按业务影响排序，不按 Excel 报错顺序

P0	日期字段存在 96 个非法或倒置值	汇总期间错误
P0	订单号重复 31 组，涉及 73 行	金额重复统计
P0	金额列混入 49 个文本值	求和不稳定
P1	客户号空值 68 个	无法关联客户
P1	科学计数法或负金额 34 个	口径待确认
P2	门店名称存在 23 组近似写法	分组不统一

四条合成记录

问题长什么样

日期不存在
2026/13/02
无法进入正确月份

日期倒置
入住 10/05 → 离店 10/03
入住时长与房晚冲突

金额混文本
1,200元
求和时被跳过或整列转文本

主键重复
DEMO-1001 × 2
可能重复统计同一订单

影响链路与排序规则

先看业务后果

影响合计、去重、期间归属的排在格式之前

再看能否确认

能指回来源和规则的，才能进入自动处理

无法回答不猜

规则缺失时保留原值，列入人工确认

字段异常

日期 / 金额 / 主键

➤

处理结果

跳过、截断或错误分组

➤

业务表象

合计差钱、订单重复、期间错位

➤

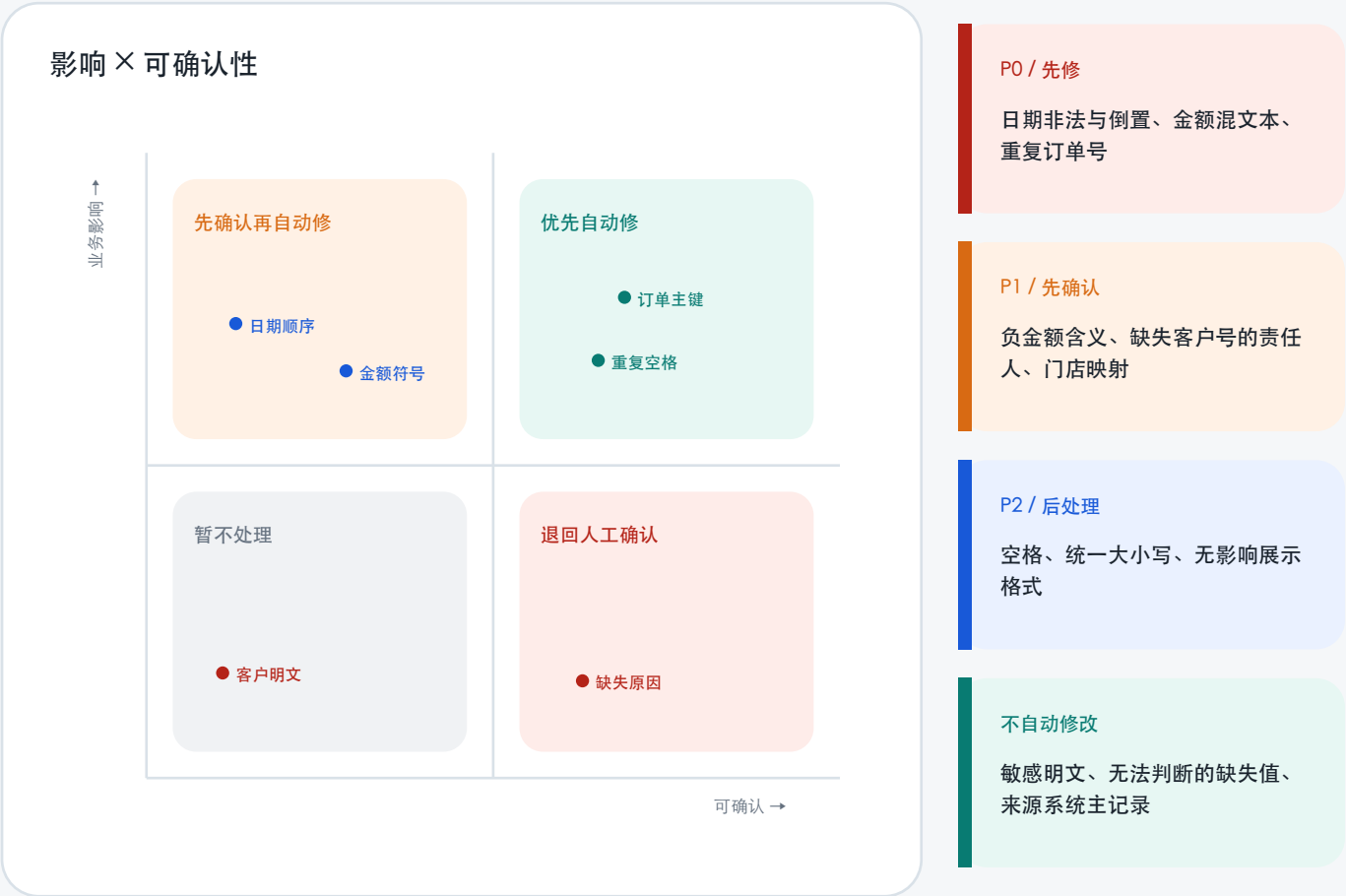
决策风险

继续自动化会把错误放大

结论：先处理会影响合计、去重和期间归属的字段；展示格式与同义值放到后一步。

不是所有错漏都应该自动改

先验的核心产出，是把“能自动处理”“必须人工确认”“应该退回来源”分开。



先验会把“猜”变成可回答的问题

真正影响交付的不是格式，而是业务规则。以下问题需要在开工前问清，不能在结果里偷偷替客户决定。

7 个规则确认项

- 01

订单号重复时，哪一条是主记录？
按更新时间、来源系统，还是人工指定？
- 02

负金额代表退款、冲销，还是录入错误？
不同含义会改变汇总与异常规则。
- 03

日期按当地时区、自然日还是业务日？
跨月与住宿天数都依赖这条规则。
- 04

客户号缺失的订单还能不能计入收入？
这决定空值是补、保留，还是退回。
- 05

金额是含税、未税还是已收？
混口径求和会产生看似正确的错误数字。
- 06

门店旧名与区域调整如何映射？
分类统一必须先有映射依据。
- 07

本次要交付的是判断、结果表，还是脚本？
三者价格和验收方式不同。

单表不能证明什么

先验报告写结论，
也写边界。

不能证明
这张表与外部系统完全一致

不能证明
缺失数据在现实中也缺失

不能证明
客户身份、经营收入或资质真实

不能代替
财务、法务或管理层的业务拍板

下一步输出

问题优先级
规则确认项
处理路线建议

05 / 下一步

199 元买到的是一份可继续决策的判断

不是无限代改，也不是先承诺一个“清洗得很漂亮”的结果。先判断，再按范围和风险选择下一步。

本样例的最终判断

先修主键、日期和金额，再做汇总。

原表建议保留。先产出问题标记与规则记录，确认后再改；如果业务方无法回答负金额和缺失客户号的归属，这两项不自动处理。

- 不覆盖原表
- 不擅填敏感值
- 不承诺结果收入

路线 A

人工按清单处理

适合一次性表、规则仍不稳定；先改 P0，再人工确认 P1。

路线 B

脚本标记与清洗

适合规则已确认、以后还会重复出现的同类表。

路线 C

先退回来源系统

适合缺失来自系统、责任归属不清或涉及敏感明文的情况。

哪些情况适合进入 199 元先验

- 一张 CSV / Excel，最多 5,000 行、30 列

● 已经脱敏，且能说明用途和期望输出

● 主要问题在空值、重复、日期、金额或列结构
- 需要逐行代改、OCR、网站抓取或系统登录

● 多文件合并、复杂业务系统建模

● 资料未脱敏，或要求保证清理、成交、收入

先免费看错漏，再判断是否值得付费

单表数据先验 · 199 元

5,000 行 / 30 列以内 · 45 分钟沟通 · 48 小时书面判断

192 5858 0147

www.igooo.com.cn/table-data-triage-199/

样例结论、字段和金额均为场景演绎；正式先验只对客户id提供并确认的脱敏输入作判断。
本样例不是客户项目、真实订单、收入或效果证明。
边界：不包含逐行代改；免费扫描器文件不会上传，付费仅收完成判断所需的脱敏样本。