

门店数据体检清单

在花钱建模之前，先花五分钟确认数据够不够

决定一个数据分析项目成败的，八成不在算法，在数据这一关。下面五个问题问完，你就知道手上的数据 能不能支撑想要的结论——任何一条答“否”，先补这一条，比直接上模型省钱得多。

#	检查项	是	否	不确定
1	粒度：数据记到多细？ 想给每家门店每个单品定备货量，数据就必须细到“门店 × 单品 × 日”。粒度不够不是清洗能解决的问题——那是信息本身不存在。 需确认：最小记录单位是什么？门店级？单品级？还是只有汇总数？	☐	☐	☐
2	跨度：覆盖了几个完整周期？ 餐饮零售季节性强。只有 6 个月数据，模型没见过春节、没见过暑期，无法把这些因素和普通波动区分开。 经验门槛：预测类至少 18 个月；选址类至少 12 个月且覆盖淡旺季。	☐	☐	☐
3	变异：你想问的那个变量，变过没有？ 想知道降价会怎样，历史上就得调过价。数据里没有变化，就没有可供学习的信号——这时该做的是设计一次小范围试验，而不是硬算。 自检方法：把问题写成“当 X 变化时，Y 会怎样”，再去数据里找 X 变化过的记录。	☐	☐	☐
4	口径：同一个字段，各部门说的是不是一回事？ “营业额”是否含税、是否扣退单、外卖佣金算哪边，财务/运营/IT 常常不一致。口径不解决，结论会在汇报会上被一句“跟财务对不上”打回来。 建议动作：要一份字段说明，拿一两个月数据和财务报表对一次账。	☐	☐	☐
5	可得性：拿数据要走多久流程？ 技术上没难度，但项目最常拖期就在这里：数据在第三方 SaaS 里、导出要审批、一次只能导一个月。 开工前敲定：谁来导、多久导一次、什么格式。	☐	☐	☐

五条都过了，接下来问自己三个问题

- 这个决策足够贵、且会重复发生吗？一次性的小决策，凭经验拍板比建模划算。
- 结论能被检验吗？预测的销量三个月后能对账、推荐的点位开业后能复盘，才叫模型；否则只是意见。
- 结论能落到具体的人和动作上吗？“优化供应链效率”没法执行；“这 12 家门店午市备货下调 15%，由区域经理在系统里改参数”可以执行。

这份清单出自我们做连锁门店选址、备货与定价项目的实际经验。体检结果有拿不准的，可以带着它找我们聊 30 分钟——**我们会直接告诉你缺什么、值不值得做，判断下来不值得做也会直说。**

智模通 Modellink（上海模视教育科技有限公司） wanghaihua@modelinsight.cn zhimotong.com