

REGISTRO DE TRATAMENTO E LIMPEZA



O que foi feito com o dado bruto, para que outra pessoa chegue ao mesmo resultado

COMO PREENCHER

- 1. Uma linha por decisão de tratamento: exclusão, preenchimento, deduplicação, conversão, junção.
- 2. Escreva o **critério**, não só a ação. "Removi outliers" não é reproduzível; "removi acima de Q3+1,5·IQR" é.
- 3. Registre **quantos registros** a decisão afetou. É o que dimensiona o risco de viés.
- 4. Guarde o dado bruto intocado. Tratamento se faz em cópia, sempre.

ANÁLISE / CÓDIGO	RESPONSÁVEL	DATA
AN-2026-014	Ana Prado, Planejamento	10/08/2026

DECISÕES DE TRATAMENTO uma linha por decisão, com critério e registros afetados

#	CAMPO / BASE	PROBLEMA ENCONTRADO	DECISÃO E CRITÉRIO	REGISTROS AFETADOS	% DA BASE	RISCO DE VIÉS
1	LEAD_TIME (Planilha Compras)	334 registros nulos de 4.180 (7,99% da base), preenchimento manual do Compras ficou em branco	Excluir enviesaria contra quem preenche pior a planilha, não quem atrasa mais; optou-se por imputar pela mediana do fornecedor.	334	7,99%	Médio, a mediana suaviza casos de lead time atipicamente alto
2	DT_PEDIDO (ERP Protheus)	53 registros com data de pedido inválida: 12 nulos, 41 fora do período analisado	Linhas excluídas: sem data de pedido válida, não dá para calcular lead time nem enquadrar no recorte de jan a jun/2026	53	0,03%	Baixo, volume irrelevante frente aos 184.320 registros
3	QTD_SALDO (WMS Vertice)	218 registros com saldo negativo por erro de lançamento, não saldo real	Saldo zerado, linha mantida: negativo vira 0 e o SKU/CD é marcado para a Logística corrigir a causa	218	0,25%	Baixo, concentrado em poucos SKUs
4	STATUS_ITEM (WMS Vertice)	1.642 registros com status fora dos 4 valores válidos (A, B, R, X), campo de texto livre	Mapeados ao valor mais próximo por tabela de equivalência (ex.: "ativo" vira A)	1.642	1,88%	Médio, o mapeamento manual introduz julgamento
5	DT_ENTREGA (Planilha Compras)	147 registros com data de entrega problemática: 51 nulos, 96 anteriores à data do pedido	51 nulos preenchidos com a mesma mediana do fornecedor usada no LEAD_TIME. Os 96 com data impossível tiveram a linha excluída	147	3,52%	Médio, mistura dois tratamentos diferentes no mesmo campo
6	FORNECEDOR (Planilha Compras)	612 registros com o fornecedor grafado de até 3 formas (ex.: "Cimentos Vale", "CIM. VALE LTDA")	Normalizados pelo cadastro do ERP Protheus, por CNPJ do fornecedor, nunca pelo nome digitado	612	14,64%	Médio, os casos ambíguos foram revisados por amostragem
7	COD_SKU × CD (junção WMS × ERP)	WMS/ERP só compartilham COD_SKU e CD, não ID de pedido: 312 pedidos com candidato ambíguo de junção	Mantido o candidato com menor diferença entre DT_PEDIDO e a movimentação do WMS; os 9 empatados foram excluídos	312	0,17%	Baixo, critério de desempate é objetivo

ANTES E DEPOIS

Base bruta: 184.320 movimentações e 4.180 pedidos com lead time. Base tratada: 184.267 movimentações (53 excluídas por DT_PEDIDO inválido) e os mesmos 4.180 pedidos, com os 334 nulos de LEAD_TIME (7,99%) preenchidos pela mediana do fornecedor. Sem o ajuste, 8% dos pedidos ficariam fora da comparação, favorecendo quem preenche melhor a planilha, não quem entrega mais rápido.