Подготовка данных: профиль, правила и журнал шагов

Файл и дефекты сгенерированы в этом прогоне; правила проверки — настоящие

1. Пригодность — это не «почистили», а «проходит проверки»

Датасет пригоден для анализа, когда по нему проходят правила: уникальность ключа, диапазон даты, знак количества, маска артикула, справочник значений. Всё остальное — вкус.

Показатель	До	После	Что сделали
Строк в файле	5047	4894	убрали дубли по ключу дата+артикул+количество+цена
Типов даты	3	1	привели к ГГГГ-ММ-ДД
Вариантов города	11	4	свели написания к 4 городам по справочнику
Пустой клиент	313	302	оставили, но пометили: клиент неизвестен — не повод терять заказ

2. Отрицательное количество не удаляют

107 строк с отрицательным количеством — это возвраты, а не мусор. Удалить их значит потерять часть выручки и не увидеть причину.

Правило	Проверка	Строк затронуто
R1 уникальность	ключ дата+артикул+количество+цена	47
R2 дата в диапазоне	01.07.2026 — 30.09.2026	5047
R3 количество > 0	иначе лист «возвраты»	107
R4 артикул по маске	SKU-###	5047
R5 город из справочника	4 допустимых значения	5047

3. Журнал важнее результата

Через месяц никто не помнит, что сделали с городом «Спб». Журнал шагов превращает чистку в воспроизводимую операцию: тот же вход даёт тот же выход.

Данные кейса сконструированы. Нормы, формулы и структура документа — настоящие.