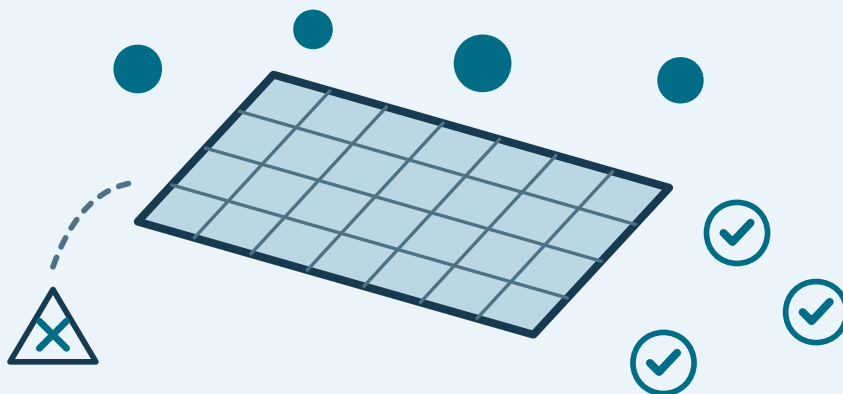




PYTHON AND DATA WORKFLOWS

Write Data Quality Rules That Explain Failures

Build dependency-aware checks with stable reasons, useful denominators and repair evidence



Sanjaay Gowda T S

FREE SAMPLE / INTRODUCTION + CHAPTER 1 / EDITION 1.0.0

sanjaaygowdats.ai

About this free sample

Write Data Quality Rules That Explain Failures

Sanjaay Gowda T S

First edition 1.0.0, 2026-10-13

Copyright 2026 Sanjaay Gowda T S. This is a labelled free sample, not the complete ebook. You may share this sample unchanged. It contains the cover, sample permissions, contents, introduction, Chapter 1 and a closing product link. The remaining chapters, worked solutions and companion files are not included. The full ebook text, diagrams and artwork have separate single-reader terms; this sample permission does not authorize sharing the full purchased PDF or EPUB.

The companion source code, tests and fictional data are separately MIT licensed. You may adapt or share that project with its included licence notice. MIT does not apply to the ebook text or book artwork. The books and companion projects have no technical activation requirement or buyer-specific watermark.

This first edition was prepared with AI assistance. The case studies are fictional teaching examples. Any measured outputs come from the supplied examples, not production customer data. The teaching scope, assumptions and limitations are explained in the chapters. Independent human technical, editorial and assistive-technology reader review is pending. Automated tests and accessibility checks do not establish certification or suitability for every use.

Run examples in a separate working folder using the supplied synthetic data. Review inputs and calculations before adapting the techniques to real operations. Model software, service access, subscriptions, ongoing updates and consulting are not included.

Author website: sanjaaygowdats.ai

Ebook terms: sanjaaygowdats.ai/ebooks/license

Inside this free sample

Start here.....	4
1. Define quality for a decision	5
Continue with the full guide	6

This free sample contains the introduction and Chapter 1 only. The closing link opens the complete guide. The companion project and worked solutions are supplied with the complete ebook.

Start here

A quality rule should help someone understand and repair a failed condition. A dashboard that merely says eighty percent good cannot explain whether the remaining records have missing dates, duplicated identities or a genuine operational delay. This book builds rules that retain their scope, dependencies and evidence instead of compressing every concern into one score.

The fictional case is Birch Parcel, which tracks small shipments on local and regional routes. All shipment identities, quantities, route limits and dates are teaching inventions. The route limits are example policy, not transport guidance or observed service performance. The program checks consistency of recorded data; it cannot establish that a parcel physically arrived or that a source record is truthful.

The companion accepts an explicit as-of date, a route configuration and shipment records. It returns record-level issues and rule-level eligible and failed counts. The supplied fixture contains one clean record, one overdue warning, one quantity and date-order failure, and one record with an unknown route and invalid planned quantity.

```
python main.py --input example-input.json
python -m unittest discover -s tests -v
```

Run the unchanged project first. Read the issue objects before their summary. Each contains a stable rule identity, severity, field, observed value and expected condition. Then inspect rule Q08: only three records are eligible for the quantity comparison, because one planned quantity is not a valid integer. Its field failure must not be counted as another failed comparison.

The nine chapters develop a rule contract, dependency gates, useful explanations and change review. Every exercise changes a meaningful boundary. The final solutions explain both the result and its limits. This is a small standard-library teaching engine, not a general schema language, monitoring service or certified data-quality framework.

1. Define quality for a decision

Start from a use, not a percentage

Birch wants a reviewer to identify shipment records that cannot support a dispatch-status report. A missing identity prevents reliable follow-up. Invalid quantities prevent meaningful comparisons. An overdue in-transit record may be accurate but require operational attention. These are different reasons for concern, so the engine separates errors from warnings.

The record grain is one shipment identity in a supplied snapshot. The source is assumed to describe current shipment status, not a history of every status transition. If the export instead contains several versions of each shipment, duplicate identity is not necessarily a source error; the grain must be revised to shipment plus revision. Define that distinction before adding a uniqueness rule.

Concern	Example	Interpretation
Validity	Planned units is text	Cannot compare quantities
Consistency	Delivery before dispatch	Recorded facts conflict
Reference	Route absent from configuration	Policy is unknown
Timeliness	Transit exceeds route limit	Review warning

No rule proves completeness of the export. Missing shipments cannot fail a row-level check because no row exists to inspect. A separate expected-population control would be needed. Similarly, a perfectly consistent record may still be false. State the boundary clearly so users do not equate rule compliance with truth.

Worked case: an overdue but valid record

S102 has valid fields and a known route. Its dispatch date is nine days before the as-of date, exceeding the fictional regional limit of four days. The engine emits a warning. Changing the date to suppress the warning would damage the data if the date is accurate. The appropriate review may be operational investigation, not source repair.

Exercise: identify the decision owner

For missing identity, impossible date order and overdue transit, name the person or role who could resolve the issue and the evidence they would need. Explain why the same automatic repair should not be applied to all three. A useful rule has a destination for its evidence, not just a place on a dashboard.

Decision worksheet: classify the consequences

Ask what would happen if each defect reached the dispatch-status report. A duplicated identity might double-count shipments. An unknown route might prevent a reviewer from applying the correct capacity policy. A date-order contradiction might undermine duration calculations. An overdue warning might identify a real delay without undermining the record's validity. Write those consequences beside the rule definitions. This makes severity a decision about use rather than a label inherited from a programming exception. The same dataset may be fit for one purpose and unfit for another: a record with an unknown route could still support a source-record count while failing a route-capacity analysis. Do not advertise a single clean-data flag as permission for every downstream use. If several consumers share the engine, publish the measured facts and let reviewed consumer policies decide which errors or warnings block their specific decisions. Preserve enough context to explain those policies later.

Continue with the full guide

This free sample ends after the introduction and Chapter 1. The complete guide contains nine teaching chapters, worked exercise solutions, primary references and a separate Python companion project. The full PDF, reflowable EPUB and project files are supplied together.

Use the complete guide to work through the remaining exercises and inspect the supplied examples. The companion code and fictional data are separately MIT licensed; the complete ebook has the single-reader terms described on its product page.

The examples are teaching fixtures, not production customer results. Independent human technical, editorial and assistive-technology reader review remains pending. Software access, subscriptions, consulting and guaranteed results are not included.

[View the complete guide and purchase options.](#)

You may share this labelled free sample unchanged. Keep the full purchased PDF and EPUB private. Thank you for reading.