

TEMPLATE FOR QAULIFIED AI SPECIALIST

Template: Data Quality Assessment & Reporting Document

Section 1: Assessment Metadata

- Dataset Name:
- Dataset Version or Time Window (if applicable):
- Assessed By:
- Date of Assessment:
- Use Case:
(e.g., predictive maintenance, fraud detection, churn analysis)
- ML Type (if applicable):
(e.g., supervised classification, unsupervised clustering)

Section 2: Data Quality Characteristic Evaluation Table

DQ Characteristic	Description	Measurement Method	Threshold / Target Value	Actual Value / Result	Conformance (✓/X)	Notes / Remediation Plan
Accuracy	Correctness of data values	Compare against verified source	≥ 95% match rate	93.2%	X	Investigate device #402 missing calibration
Completeness	Proportion of required data present	% of non-null critical fields	≥ 98%	99.5%	✓	

Timeliness	Speed of data availability	Avg. ingestion delay (sec)	≤ 5 seconds	7.8 seconds	X	Real-time stream delay during peak hours
Representativeness	Coverage of expected population	Compare class or region counts	All regions/seasons present	Missing rural data Q1	X	Augment with external dataset
Consistency	Adherence to format and schema	Schema validation report	100% schema-conformant	100%	✓	

Section 3: Visual Summary (Optional – for Excel or Word)

You can insert:

- Bar chart: Compliance rate by DQ characteristic
- Heatmap: High/Medium/Low risk indicators
- Line graph: Trend of data quality over time (for continuous learning datasets)

Section 4: Overall Dataset Assessment Summary

- Does the dataset meet the minimum quality requirements for intended use?
☐ Yes ☐ No ☐ Partially
- Key Quality Issues Identified:
 1. Ingest delay exceeds threshold during peak streaming
 2. Underrepresentation of rural region data
 3. Slight drop in accuracy in sensor batch #2



- Recommended Actions:
 - Normalize ingestion rates with upgraded stream pipeline
 - Source and incorporate data from missing geographic region
 - Recalibrate sensor batch #2 and update ingestion logs
- Impact Level on ML Model Performance (if known):
 - ☐ None ☐ Low ☐ Medium ☐ High

Explanation: Representativeness and accuracy issues may reduce prediction reliability for rural regions.

📁 Section 5: Governance and Audit

- Data Owner / Steward:
- Reviewer / QA Approver:
- Date of Next Planned Reassessment:
- DQ Assessment Version #: