



TEMPLATE FOR QAULIFIED AI SPECIALIST

Template: Data Quality Requirements Document

Section 1: Project Context

- Project Name:
- Date Created:
- Author/Team:
- Reviewed By:
- Intended Use Case (Analytics or ML Type):
(e.g., churn prediction, fraud detection, demand forecasting)
- Key Stakeholders:
(e.g., data scientists, data stewards, ML engineers, compliance team)

Section 2: Data Sources and Description

Data Source Name	Description	Internal/External	Data Format	Collection Frequency
-----------------------------	--------------------	--------------------------	------------------------	---------------------------------

Section 3: Data Usage Requirements

- Data Purpose:
(What problem will this data support solving?)
- Expected Input Types:
(e.g., sensor logs, customer profiles, text, audio)
- Expected Output Type:
(e.g., classification label, regression score, recommendation)
- Performance Objectives:
(e.g., >90% accuracy, <5% error rate)

Section 4: Data Quality Characteristics & Thresholds

DQ Characteristic	Description	Threshold / Minimum Requirement	Priority (High/Med/Low)	Notes
Accuracy	Data values correctly represent real-world conditions	$\geq 95\%$ accuracy in location data	High	GPS precision critical
Completeness	No missing or null values in critical fields	$<2\%$ missing values allowed	High	Critical for ML learning
Timeliness	Data is available when needed	≤ 5 sec delay in stream ingestion	Medium	Applies to real-time data
Representativeness	Data distribution reflects production population	All regions & seasons included	High	Avoids model bias
Consistency	Data follows a standard format or structure	Valid JSON/XML or schema aligned	Medium	
Uniqueness	No duplicate entries or IDs	100% unique identifiers	Medium	

Section 5: Risk of Not Meeting Requirements

DQ Characteristic	Business/Model Impact if Not Met	Mitigation Approach
Accuracy	Misleading predictions, poor decisions	Quality audits; cross-check with labels



Completeness	Model underperformance, training interruptions	Imputation; outlier detection
Representativeness	Biased models, ethical risks, regulatory noncompliance	Data resampling; enrichment strategies

Section 6: Evaluation and Measurement Plan

- Tools to be used:
(e.g., Python scripts, data profiling tools, Great Expectations)
- Frequency of DQ Checks:
(e.g., Daily, Per pipeline run, Weekly)
- Roles & Responsibilities:

Role	Name	Responsibility
Data Steward		Monitor DQ metrics and escalate deviations
ML Engineer		Validate model performance vs DQ criteria
Compliance Lead		Ensure ethical/representative data use

Section 7: Approval and Review

- Reviewed by:
- Date Reviewed:
- Approved by:
- Revision Schedule: