Data Quality
Assessment Tool

Type: Assessment Tool

Target Audience: Data Engineers, ML Engineers, Data Governance Teams

This tool evaluates datasets across six dimensions to determine their fitness for training AI models. Score each dimension and calculate overall data quality readiness.

Assessment Information

Field	Details
Dataset Name	
Assessment Date	
Assessor	
Intended AI Use Case	
Data Source(s)	

Dimension 1: Completeness

Measures the extent to which required data is present.

Metric	Measurement	Score (1-5)
Missing Value Rate	% of null/empty values across all fields	
Required Fields Population	% of mandatory fields populated	
Record Completeness	% of records with all required fields	
Coverage	Data available for all expected categories/time periods	

Completeness Score: _____ / 5

Dimension 2: Accuracy

Measures how well data reflects real-world truth.

Metric	Measurement	Score (1-5)
Error Rate	% of values known to be incorrect	
Validation Rate	% validated against authoritative source	
Outlier Analysis	Proportion of statistically anomalous values	
Label Accuracy	For supervised learning: label correctness rate	

Accuracy Score: _____ / 5

Dimension 3: Consistency

Measures uniformity of data formats and logical coherence.

Metric	Measurement	Score (1-5)
Format Standardization	Adherence to defined data formats	
Cross-Field Logic	Logical consistency between related fields	
Duplicate Rate	% of duplicate records	
Referential Integrity	Validity of foreign key relationships	

Consistency Score: _____ / 5

Dimension 4: Timeliness

Measures how current and frequently updated the data is.

Metric	Measurement	Score (1-5)
Data Age	Time since data collection/last update	
Refresh Frequency	How often data is updated	
Latency	Delay between event and data availability	
Currency	Data reflects current state of the world	

Timeliness Score: _____ / 5

Dimension 5: Relevance

Measures alignment with AI use case requirements.

Metric	Measurement	Score (1-5)
Feature Coverage	Required features present in dataset	
Use Case Alignment	Data matches intended AI application	
Signal-to-Noise	Proportion of useful vs. irrelevant data	
Granularity	Level of detail appropriate for use case	

Relevance Score: _____ / 5

Dimension 6: Representativeness

Measures how well training data matches the target population.

Metric	Measurement	Score (1-5)
Population Coverage	All target segments represented	
Distribution Match	Training data mirrors production distribution	
Demographic Balance	Protected groups adequately represented	
Edge Case Coverage	Rare but important cases included	

Representativeness Score: _____ / 5

Overall Data Quality Summary

Dimension	Score	Weight	Weighted Score
Completeness	___/5	20%	
Accuracy	___/5	25%	
Consistency	___/5	15%	
Timeliness	___/5	15%	
Relevance	___/5	15%	
Representativeness	___/5	10%	
TOTAL		**100%**	**___/5**

Readiness Thresholds

Score	Status	Recommendation
4.0 - 5.0	Ready	Data suitable for AI model training
3.0 - 3.9	Conditional	Address gaps before production use
< 3.0	Not Ready	Significant remediation required