

Dataset Readiness Checklist & Quality Comparison

Ensuring Data Quality Before Modeling

Dataset Readiness Checklist

Use this checklist before you start work. Confirm each item. Record gaps and actions.

Check Item	Yes/No	Notes	Action Needed
Data has a clear source			
Data has a clear purpose			
Columns have clear meanings			
Units of measure are consistent			
Dates use one format			
No duplicate rows			
Missing values are identified			
Outliers are reviewed			
Labels match the task			
Sensitive data is removed			
Data reflects the real world			
Data covers the right time range			
Class balance is understood			
Sampling method is documented			
Data ownership is clear			

Readiness Summary

Describe the current state in one short paragraph.

Strengths:

Weaknesses:

Next Actions:

Good vs Weak Data Comparison

Use this sheet to judge quality. Compare examples. Write what each case implies.

lightgray Dimension	Good Data	Weak Data	Impact on Results
Clarity	Columns have clear names	Columns have vague names	
Consistency	Values follow one format	Values mix formats	
Coverage	Data spans relevant time	Data misses key periods	
Accuracy	Values match reality	Values contain errors	
Completeness	Few missing values	Many missing values	
Balance	Classes are well represented	One class dominates	
Freshness	Data reflects current conditions	Data is outdated	
Relevance	Features support the task	Extra noise in features	
Documentation	Source and method are known	No context provided	
Ethics	Sensitive fields removed	Private data included	

Reflection

Main strengths in this dataset:

Main risks in this dataset:

Changes required before modeling: