Data Collection and Preprocessing Phase

Date	1 August 2025
Skillwallet ID	SWUID20250194750
Project Title	Anemia Sense: Leveraging Machine Learning For Precise Anemia
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in 'Hemoglobin', 'MCHC', 'MCV', and 'MCH' columns	High	Apply mean or median imputation, or use clinically validated ranges for plausible medical values.
Kaggle Dataset	Categorical data in the 'Gender' and 'Result' columns	Moderate	Apply label encoding or one-hot encoding where appropriate.