Data Collection and Preprocessing Phase

Date	8 August 2025
Skillwallet ID	SWUID20250188620
Project Title	Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognition
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report summarizes issues found in the selected anemia dataset, their severity, and the resolution plans. This process ensures the dataset is consistent, accurate, and ready for model training.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
anemia.csv	Gender column stored as categorical values (Male/Female)	Moderate	Applied label encoding (Male=1, Female=0)
anemia.csv	Outliers in Hemoglobin and MCV values	Low	Retained medically valid extreme values after verification
anemia.csv	No missing values in dataset	None	No action needed