

Data Collection and Preprocessing Phase

Date	05 July 2025
Team ID	SWTID1749835773
Project Title	Applicant Credibility Prediction For Loan Approva
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: Loan_ID 0 Gender 13 Married 3 Dependents 15 Education 0 Self_Employed 32 ApplicantIncome 0 CoapplicantIncome 0 LoanAmount 22 Loan_Amount_Term 14 Credit_History 50 Property_Area 0 Loan_Status 0	Moderate	Use mean/median imputation.

Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data.
----------------	---------------------------------	----------	--------------------------------------