

Data Collection and Preprocessing Phase

Date	18 June 2025
Team ID	XXXXXX
Project Title	sloan digital sky survey (sdss) galaxy classification using machine learning
Maximum Marks	2 Marks

Data Quality Report

The Data Quality Report will highlight data quality concerns identified in the chosen source, detailing their severity and proposed solutions. It will serve as a structured approach to detect and resolve data inconsistencies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in the u, g, r, i, z, petroR50_u, petroR50_g, petroR50_i, petroR50_r, petroR50_z, psfMag_u, psfMag_r, psfMag_g, psfMag_i, psfMag_z, expAB_u, expAB_g, expAB_r, expAB_i, expAB_z	Moderate	Removed rows with missing values
Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data