

Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720174957
Project Title	Human resource management employee promotion prediction
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template 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 Source	Data Quality Issue	Severity	Resolution Plan
Employee Records	There are some null values	Moderate	By using is.na we have resolved the solution
Employee Records	Missing values in key fields	High	Implement data imputation methods such as mean/mode imputation for numerical and categorical data, or use advanced techniques like KNN imputation.

Historical Promotion Data	Outliers and anomalies in promotion records (e.g., unusually rapid	High	Utilizing statistical techniques or subject-matter expertise, locate and look into anomalies. Based on
---------------------------	--	------	--

	promotions)		validation with HR specialists and consistency checks, eliminate or fix anomalies. Use a strong outlier detection algorithms like IQR or Z-score methods
Employee Surveys	Duplicate survey responses	Moderate	Make use of data deduplication strategies, such as timestamp analysis and unique employee identification, to find and eliminate duplicate rows. Make sure every worker only submits one survey response.
Skills and Certifications	Outdated or unverified skills and certification data	Moderate	Update your qualification and skill records on a regular basis by using self-reporting tools and verifying your information with professional certifying organizations. To guarantee data accuracy, implement routine audits and verification procedures.