Data Collection and Preprocessing Phase

Date	27 June2025
Team ID	NONE
Project Title	Employee Performance Prediction using Machine Learning

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 the wip (Work in Progress) column	Moderate	Imputed using the median value, as the distribution is skewed and mean would be sensitive to outliers
Kaggle Dataset	Categorical variables: quarter and department	High	Label encoded using LabelEncoder to convert into numerical format suitable for machine learning models
Kaggle Dataset	Inconsistent spacing in department values (e.g., "finishing " with trailing space)	Low	Cleaned using .strip() to remove leading/trailing whitespaces for consistency