

Data Collection and Preprocessing Phase

Date	05 August 2025
Skillwallet ID	SWUID20250186419
Project Title	Employee Performance 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
<i>garments_worker_productivity.csv</i>	Missing values in the <i>wip</i> (Work In Progress) column. Based on the sample data, at least one row has a null value.	Moderate	Fill the missing values with the mean of the <i>wip</i> column. This is a simple and effective method for handling a small number of missing values in a numerical column.
<i>garments_worker_productivity.csv</i>	Categorical features (<i>department</i> , <i>quarter</i> , <i>day</i>) are not in a numerical format suitable for machine learning models.	Low	Apply one-hot encoding to nominal categorical features like <i>department</i> and <i>quarter</i> . Apply label encoding or similar techniques to other categorical features as needed, such as the <i>day</i> column to prepare it for use.
<i>garments_worker_productivity.csv</i>	The <i>date</i> column is in a string/object format, preventing direct use in correlation analysis or as a time-series feature.	Low	Convert the date column to a datetime object. Extract new, useful numerical features such as <i>day_of_week</i> and <i>month</i> from this column. After feature

			engineering, the original date column will be dropped.
garments_ worker_pro ductivity.cs v	Potential outliers or anomalies may exist in numerical columns such as smv, wip, over_time, and incentive. These can negatively impact model performance.	Moderate	Use statistical methods and visualizations (e.g., box plots) to detect outliers. Depending on the nature and count of the outliers, they will either be removed, capped (winsorized), or transformed to minimize their impact on the model.