Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	SWTID1720108643
Project Title	Garment worker productivity prediction
Maximum Marks	2 Marks

Data Quality Report Template

To ensure data-driven decision making, a robust Data Quality Report Template has been developed. This template streamlines the identification and rectification of data discrepancies within our chosen data source. By capturing data quality issues, their severity levels, and corresponding resolution plans, the template facilitates a systematic approach to improving data accuracy and overall project reliability.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in WIP (Finishing department)	Medium	Impute missing values with 0 for Finishing department. Explore reasons for missing data and potentially implement data collection improvements.
Kaggle Dataset	Frequent value of 0 in Idle_men, idle_time, no_of_style_change	Low	Investigate reasons behind frequent zeros. Consider adjusting data collection process to capture more granular information (e.g., dropdown menus instead of open-ended fields for these values).