

Data Collection and Preprocessing Phase

Date	4 July 2024
Team ID	SWTID1720090524
Project Title	Garment Worker Productivity 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
Kaggle Dataset	Missing values in 'work_in_progress' column.	Low	<p>Filling in missing values with median as opposed to mean since the data is skewed.</p> <pre>[] #checking skewness df.work_in_progress.skew() 9.741786273952965 [] #filling in missing values with median as opposed to mean since the data is skewed df['work_in_progress'].fillna(df['work_in_progress'].median(),inplace=True)</pre>
Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data.