

Data Collection and Preprocessing Phase

Date	5 May 2024
Team ID	737906
Project Title	Walmart Sales Analysis For Retail Industry
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
Dataset-1	Missing Data	Low	<p>The dataset may contain missing values, which can pose challenges during analysis and modeling. Addressing missing data requires careful consideration, including imputation techniques or removing observations with missing values.</p> <p>Imputation: Replace outliers with a measure of central tendency (e.g., mean, median) to maintain the overall distribution of the data. Use interpolation techniques to</p>

			estimate outlier values based on neighboring data points.
Dataset-1	Outliers and Anomalies	moderate to high	<p>Transformation: Apply transformations like log transformation to stabilize variance and reduce the impact of outliers.</p> <p>Removal: Remove outliers from the dataset entirely if they are deemed to be erroneous or unrepresentative of the underlying distribution.</p>