

Data Collection and Preprocessing Phase

Date	07 April 2024
Team ID	722312
Project Title	Walmart Sales Analysis For Retail Industry With Machine Learning
Maximum Marks	2 Marks

Data Quality Report

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
Walmart Sales Dataset	Mention the issues faced in the selected dataset.	High	Implement imputation techniques like mean/median for numerical features and mode for categorical features. Alternatively, consider removing rows/columns with excessive missing values.

Walmart Sales Dataset	Outliers	Moderate	Detect outliers using z-score or IQR and handle them by clipping, winsorizing, or removing. Alternatively, use robust statistical measures or transformation techniques.
Walmart Sales Dataset	Inconsistent Data Formats	Moderate	Standardize date/time formats, ensure consistency in currency formats, and units of measurement.
Walmart Sales Dataset	Duplicate Records	Low	Identify and remove duplicate records based on key identifiers. Alternatively, consolidate information or merge duplicate records if they represent different instances.
Walmart Sales Dataset	Data Inconsistencies	Moderate	Validate data against predefined business rules or constraints. Cleanse or correct inconsistent data based on domain knowledge or data profiling.