

Data Collection and Preprocessing Phase

Date	20 July 2024
Team ID	SWTID1720110595
Project Title	Ecommerce Shipping Prediction Using Machine Learning
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
E-commerce Shipping Dataset	Missing values	Moderate	Use forward fill method to impute missing values
E-commerce Shipping Dataset	Inconsistent data types	Moderate	Convert data types to appropriate types (e.g., integers, floats, categories). <code>data['Column'] = data['Column'].astype('int')</code>
E-commerce	Outliers in numerical data	High	Use Z-score to identify and remove outliers. <code>data_cleaned =</code>

Shipping Dataset			<code>data[(np.abs(stats.zscore(data[numerical_features])) < 3).all(axis=1)]</code>
E-commerce Shipping Dataset	Unnecessary columns	Low	Drop columns that are not useful for analysis. <code>data.drop(columns=['Unnecessary_Column'])</code>
E-commerce Shipping Dataset	Lack of normalization in numerical features	Moderate	Apply standard scaling or normalization techniques. <code>data[numerical_features] = StandardScaler().fit_transform(data[numerical_features])</code>
E-commerce Shipping Dataset	Incorrect values due to data entry errors	Moderate	Identify and correct or remove erroneous data through validation rules or manual inspection. <code>data = data[data['Column'] <= valid_range]</code>