Data Collection and Preprocessing Phase

Date	05 May 2024
Team ID	team-739671
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 Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data
Kaggle Dataset	Handling outliers in data	Moderate	IQR is used

Data Source: Customer Order Database

Data Source	Data Quality Issue	Severity	Resolution Plan
Customer Order Table	Missing values in the 'OrderDate' field..	High	Implement a mandatory check for 'OrderDate' field before data entry

	Inconsistent formatting in the 'ID' field (e.g., mix of numbers and letters).	Medium	Apply a standard format and use a script to reformat existing entries.
	Outdated 'ShippingAddress' information.	Low	Regularly update addresses through automated address validation APIs.

Data Source: Sales Performance Dataset

Data Source	Data Quality Issue	Severity	Resolution Plan
Sales Performance Table	Incorrect 'SalesAmount' values due to manual entry errors.	High	Implement validation rules and automate data entry where possible.
	Missing 'SalesPerson' identifiers in some records.	Medium	Perform a data audit to identify missing values and rectify them.
	Redundant columns with no relevant data.	Low	Conduct a data review to identify and remove redundant columns.
	'Region' field has ambiguous names leading to misclassification.	Medium	Standardize region names and ensure consistent use across all datasets.