

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720161281
Project Title	Ecommerce Shipping Prediction Using Machine Learning
Maximum Marks	2 Marks

Data Quality Report Template

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Data is imbalanced. The data for the scenario when product is received on time is more than not received on time	Moderate	Data handling . SMOTE technique .
Kaggle Dataset	Data contains string values like mode of shipment-Flight , Ship ,etc.	High	Data Encoding Techniques. Label Encoder.
Kaggle Dataset	Data variables contains different values small and large which can induce bias in ML model.	High	Data scaling techniques. Standard Scaler.