

## Data Collection and Preprocessing Phase

Date	8 JULY 2024
Team ID	SWTID1720108776
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.

<b>Data Source</b>	<b>Data Quality Issue</b>	<b>Severity</b>	<b>Resolution Plan</b>
Source Name: Kaggle	It is a publicly available dataset in Kaggle of size 124KB,  The dataset contains many Features to predict on, and also have many outliers in two columns (Features), and data is imbalance on target Feature because the number of 1's, 0's are different. The dataset has 11000 Rows which is huge.	Moderate	We are going, Find the outliers and replace them with Mean value of the column, because removing them causing 2000 data points loss and remaining 8 columns (Features) which are valuable for prediction are removing because of just 2 columns.  Data Imbalance is to be solved by using SMOTE over sampling.