

Data Collection and Preprocessing Phase

Date	09 July 2024
Team ID	SWTID1720084775
Project Title	Ecommerce Shipping Prediction using Machine Learning
Maximum Marks	2 Marks

Data Quality Report

The Data Quality Report summarize data quality issues from the selected source <https://www.kaggle.com/datasets/prachi13/customer-analytics>, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
https://www.kaggle.com/datasets/prachi13/customer-analytics	<ol style="list-style-type: none"> there's not too much variation between determinants and their results. "Reached.on.Time _ Y.N" feature was unbalanced with 1 being a predominant outcome 	Low	<ol style="list-style-type: none"> Analyze the existing determinants and results in the dataset and identify if there are any additional features or derived variables that can be created. Feature engineering involves transforming or combining existing features to create new ones that may capture more information or patterns in the data. This can help introduce more variation and improve the predictive power of your models.

			<p>2. We employed undersampling to mitigate bias in the dataset. This approach involves reducing the number of instances in the majority class to ensure a more balanced distribution among the two different classes. By doing so, we can enhance the model's ability to learn from and accurately predict the minority class, leading to improved overall performance and reduced bias in the results</p>
--	--	--	---