

Data Collection and Preprocessing Phase

Date	03 October 2024
Team ID	LTVIP2024TMID24974
Project Title	Analysis Of Amazon Cell Phone Reviews
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report outlines data quality issues from the selected Amazon cell phone reviews dataset, including their severity levels and the proposed resolution plans. This report helps systematically identify and address any discrepancies or issues with the dataset to ensure high-quality input for the machine learning model.

Data Source	Data Quality Issue	Severity	Resolution Plan
Amazon Cell Phone Reviews Dataset	Duplicate reviews based on "Review ID" and "Review Text"	Moderate	Identify and remove duplicate entries to ensure uniqueness in reviews.
Amazon Cell Phone Reviews Dataset	Star rating skew towards 5 stars (rating imbalance)	Moderate	Apply weighting or filtering to account for bias and provide a balanced analysis
Amazon Cell Phone Reviews Dataset	Outliers in "Helpful Votes" field (unusually high counts)	High	Investigate and flag for potential manipulation or spam, potentially removing those reviews.