

Data Collection and Preprocessing Phase

Date	14 June 2024
Team ID	-
Project Title	Golden Harvesting: A Predictive Model For Apple Quality Assurance
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 Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in columns.	Low	Drop rows with missing value.
Kaggle Dataset	Invalid value in A_id	Low	Drop the row with invalid A_id value.
Kaggle Dataset	Outliers in Size	Medium	Cap the outliers to the determined whisker values.
Kaggle Dataset	Outliers in Weight	Medium	Cap the outliers to the determined whisker values.

Kaggle Dataset	Outliers in Sweetness	Medium	Cap the outliers to the determined whisker values.
Kaggle Dataset	Outliers in Crunchiness	Medium	Cap the outliers to the determined whisker values.
Kaggle Dataset	Outliers in Juiciness	Medium	Cap the outliers to the determined whisker values.
Kaggle Dataset	Outliers in Ripeness	Medium	Cap the outliers to the determined whisker values.
Kaggle Dataset	Outliers and non-numeric Acidity values	High	Convert 'Acidity' to numeric, coercing errors, and drop any resulting NaN values.