

## Data Collection and Preprocessing Phase

Date	4 Oct 2025
Team ID	SWUID20250181666
Project Title	Global Food Production Trends and Analysis: A Comprehensive Study from 1961 to 2023 Using Power BI
Maximum Marks	3 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 production columns for certain years and regions	Moderate	Use Power Query to filter nulls, apply interpolation or impute with regional averages
Kaggle Dataset	Duplicate entries for country-commodity combinations	Low	Remove duplicates using Power BI's "Remove Duplicates" function in Power Query
Kaggle Dataset	Inconsistent country codes and naming conventions	Moderate	Standardize country names using a reference mapping table and apply transformations

Kaggle Dataset	Text stored in numeric fields (e.g., “N/A” in production volume)	High	Replace non-numeric text with nulls, convert column to numeric type, validate ranges
Kaggle Dataset	Misaligned columns due to formatting issues in raw CSV	Moderate	Re-import using correct delimiter settings, manually inspect and realign columns
Kaggle Dataset	Outliers in production data (e.g., sudden spikes or drops)	Moderate	Use statistical methods (Z-score, IQR) to detect and flag; validate against external sources