

Data Collection and Preprocessing Phase

Date	5 may 2025- 6 May 2025
Team ID	xxxxxxx
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
Dataset	Missing values for food production in certain years or countries.	Moderate	Use Power Query to identify null values and apply imputation methods (e.g., mean substitution or interpolation). For critical missing data, flag and exclude from analysis.
Dataset	Inconsistent units High of measurement (e.g., tonnes vS. kilograms).	High	Standardize all measurement units using Power Query transformations to convert values

			into a consistent unit (e.g., metric tonnes).
Dataset	Duplicate rows for the same country-year-crop combination.	Low	Use Power BI's 'Remove Duplicates' feature in Power Query Editor based on key fields (Country, Year, Crop).
Dataset	Incorrect country names or inconsistent spellings (e.g., 'United States', 'USA').	Moderate	Normalize country names using a lookup table or replace values using Power Query's Replace function.
Dataset	Outliers in production values. (extremely high or	Moderate	Identify outliers using statistical thresholds (e.g. Z-scores or IQR) and review data sources. Apply transformations or cap values where appropriate.