

Data Collection and Preprocessing Phase

Date	30 June 2025
SkillWallet ID	SWUID20250170810
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 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 – World Food Production Dataset	Missing values in the 'Value' (Production in tonnes) column for some records	High	Filter out or impute missing records depending on context. For major crops with many missing values, apply average or median by crop & region using Power Query or DAX.
Kaggle – World Food Production Dataset	Duplicate rows present with the same Year, Item, Entity, and Value	Moderate	Use Power Query → Remove Duplicates based on relevant key columns: Year, Item, Entity.
Kaggle – World Food Production Dataset	Inconsistent naming of entities (e.g., "United States" vs. "United States of America")	Moderate	Standardize names using a mapping table or Power Query's "Replace Values" feature.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle – World Food Production Dataset	Some entries contain 0 production values, likely due to missing or misreported data	Moderate	Filter out zero-value rows only if they are outliers or not representative; otherwise, retain them with annotations.
Kaggle – World Food Production Dataset	Column datatypes not properly defined (e.g., Year as text, Value as text)	Low	Convert 'Year' to whole number and 'Value' to decimal number using Power Query or Power BI data model settings.
Kaggle – World Food Production Dataset	Redundant or irrelevant crop items (e.g., combined categories or outdated classifications)	Low	Filter and clean crop names; optionally group related items into broader categories for clearer insights.