# Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 30 July 2025 |
| ID | SWUID20250188226 |
| Project Title | Global Food Production Analysis (1961-2023) |
| Maximum Marks | 10 Marks |

## Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| Section | Description |
| Data Overview | This dataset contains information on global food production including various fruits (e.g., apples, bananas, grapes, oranges) and staple crops (e.g., wheat, rice, potatoes). The dataset includes country-wise and year-wise production volumes measured in metric tons. |
| Data Cleaning | Checked for missing values and duplicates. Handled missing numeric entries by replacing them with median values. Removed exact duplicate rows if any. |
| Data Transformation | Used Power Query to filter fruits and crops separately, sort by production volume, and create new calculated columns like 'Production in Million Tons'. In this section we also change column from one datatype to another such as from decimal to whole number. |
| Data Type Conversion | Converted all production values stored as text to appropriate numeric datatypes. Verified consistency of country names and categories. |