Data Collection and Preprocessing Phase

Date	13 September 2025
Team ID	SWUID20250172428
Project Title	Global Food Production Trends and Analysis: A Comprehensive Study from 1961 to 2023 Using Power BI
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	The dataset is sourced from FAOSTAT and Kaggle, covering food production from 1961–2023. It includes attributes like year, crop, area harvested, production, and yield in structured CSV/Excel format.
Data Cleaning	Missing values were handled using averages or interpolation. Duplicate rows were removed, and inconsistencies in names (countries/crops) were standardized for accuracy.
Data Transformation	Power Query was used to filter irrelevant attributes, sort data by year and country, create calculated columns (e.g., Yield = Production ÷ Area), and pivot tables for easier visualization.
Data Type Conversion	Converted Year to whole number, numeric columns (Production, Yield, Area) to decimals, and Country/Crop names to text to ensure proper analysis.
Column Splitting and Merging	Split combined fields (e.g., “1961F” → Year and Type) and merged key columns (e.g., Country + Crop) to create unique identifiers.
Data Modeling	Built relationships between fact (production) and dimension (year, crop, country) tables. Created DAX measures like Total

	Production, Total Area, and Avg Yield for analysis.
Save Processed Data	Final cleaned dataset and model were saved in Power BI (.pbix). Raw and processed data backups were also maintained for future use.