Data Collection and Preprocessing Phase

Date	2 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	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 includes annual production data (in tonnes) for crops like rice, wheat, maize, apples, bananas, tea, coffee, and more across countries and regions from 1961 to 2023. It serves as the foundation for identifying long-term agricultural trends and regional contributions.
Data Cleaning	Removed missing values, handled duplicates, corrected inconsistent country codes, and standardized column names for clarity. This ensures the dataset is reliable and free from anomalies that could distort insights.
Data Transformation	Used Power Query to filter by year and commodity, sort by production volume, pivot data for comparative analysis, and create calculated columns for growth rates and regional totals. These transformations enable flexible slicing and dicing of data for deeper exploration.
Data Type Conversion	Ensured numeric fields (e.g., production volumes) were correctly typed as integers or decimals; converted year and entity codes to appropriate formats. Accurate data types are essential for correct aggregations and visualizations.
Column Splitting and Merging	Split composite columns (e.g., “Entity–Code”) and merged crop categories where needed to simplify analysis and improve dashboard usability. This step enhances readability and aligns the structure with analytical goals.

Data Modeling	Defined relationships between fact tables (production data) and dimension tables (entities, years, commodities); created DAX measures like total production. A strong data model ensures efficient querying and dynamic reporting.
Save Processed Data	Saved the cleaned and modeled dataset in Power BI (.pbix format); exported key tables for reuse and documentation in future projects. This allows for version control, reproducibility, and easy sharing with stakeholders.