Data Collection and Preprocessing Phase

Date	3 October 2025
Team ID	xxxxxx
Project Title	Global Food Production Trends and Analysis 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 consists of global production values for crops including rice, wheat, maize, coffee, tea, grapes, apples, bananas, and oranges, measured annually from 1961 to 2023 for most countries and regions. The data is sourced from the FAO and includes crop output volumes, entities, and temporal coverage. Initial exploration included checking the source format (CSV/Excel), the range of crops and regions, and identifying missing or inconsistent records.
Data Cleaning	Handled missing values for specific years by interpolation, ensuring continuity in time series analysis
Data Transformation	Used Power Query to filter relevant years and crops, sort data by year and entity, pivot data to wide format, and created calculated columns for production totals and growth percentages.
Data Type Conversion	Converted year columns to whole number (date) format, production values to numeric (decimal) types for accurate calculations, and entity names to text fields for consistent filtering and grouping in Power BI.

Column Splitting and Merging	-
Data Modeling	Defined relationships between core tables—such as crop production, country/entity, and year—inside Power BI, enabling flexible analysis and supporting measures like annual total production, growth rate, and crop share.
Save Processed Data	Saved the cleaned and transformed dataset as a Power BI (PBIX) data model for dashboard development and exported a cleaned CSV file for reproducibility and future use in other tools.