

Data Collection and Preprocessing Phase

Date	6 October 2025
Team ID	-----
Project Title	Global Food Production Trends And 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	Summarizes scope, sources, units, and geographic coverage, outlining key variables like crop categories, regions, and annual production totals.
Data Cleaning	Standardizes country and commodity names, resolves duplicates, fills or flags missing yearly entries, and corrects obvious outliers.
Data Transformation	Builds analysis-ready fields such as year-over-year growth, per-capita output, regional aggregates, and inflation/PPP-adjusted measures where applicable.
Data Type Conversion	Ensures consistent numeric, date, and categorical types for efficient filtering, modeling, and time-series operations.
Column Splitting and Merging	Separates composite labels (e.g., “Commodity – Subtype”) and merges granular categories into comparable groups across countries and decades.
Data Modeling	Establishes star-schema style relationships among Facts (production by year, commodity, country) and Dimensions (time, geography, commodity taxonomy), with calculated measures for trends and indices.

Save Processed Data	Exports validated, versioned datasets and metadata dictionaries for reproducible analysis and dashboarding.
---------------------	---