3. Data Collection and Preprocessing Phase

Date	03 Oct 2025
Team ID	xxxxxx
Project Title	Analysis and Visualization of Global Food Production Data (1961–2023)
Maximum Marks	10 Marks

3.1. Data Exploration and Preprocessing

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 contains **11,912 rows and 24 columns**, covering **global food production trends from 1961 to 2023**. It includes multiple commodities such as cereals (rice, wheat, maize), fruits (grapes, apples, bananas, oranges, avocados), and other agricultural products (tea, coffee, cocoa, yams, potatoes, chicken meat, palm oil, etc.). • **Entity:** Country/Region name • **Year:** 1961–2023 • **Production columns:** 22 commodities measured in **tonnes**
Data Cleaning	**Missing Values:** Checked for null entries; dataset was found to have complete records. **Duplicates:** No duplicate rows detected for (Entity, Year) pairs. **Outliers:** Extreme values were detected in certain production volumes. Outlier handling was performed

	by: • Validating unusually high values against global averages. • Applying **visual inspections in Power BI** and removing/adjusting anomalies. **Error Corrections:** Minor inconsistencies in column naming were standardized (spacing in column names).
Data Transformation	**Power Query** in Power BI was used for: • **Filtering** unnecessary attributes. • **Sorting** commodities by year/production
Data Type Conversion	• Converted all **production columns** to numerical data type (**Whole number**) for accurate aggregation. • Converted **Year** to **date/time hierarchy** for time-series analysis in Power BI. • Ensured **Entity** is stored as categorical (text) for grouping.
Column Splitting and Merging	• Standardized column names by removing extra spaces (e.g.,Rice Production (tonnes) → Rice Production (tonnes)). • Merged similar category fields for uniform visualization.
Save Processed Data	• Final cleaned dataset was saved in **Power BI Data Model**. • Processed file also exported as **CSV** for backup.