Data Collection and Preprocessing Phase

Date	25 – 12 - 2025
Team ID	xxxxxx
Project Title	Global Food Production Analysis (1961–2023)
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

This phase focuses on exploring the dataset, identifying data quality issues, and applying preprocessing techniques to ensure accuracy, consistency, and readiness for visualization and analysis using Power BI.

Section	Description
Data Overview	The dataset contains historical global food production data from 1961 to 2023. It includes information on crop types, production quantities (in tonnes), years, and entities such as countries or regions. The dataset supports analysis of long-term trends, regional contributions, and crop-wise production comparisons.
Data Cleaning	Missing values were identified in certain crops and years and handled using Power Query by either removing incomplete rows or replacing values where appropriate. Duplicate records were detected and removed to ensure data accuracy. Inconsistent naming conventions were standardized to maintain uniformity.
Data Transformation	Power Query was used extensively for filtering irrelevant rows, sorting data by year and crop type, pivoting and unpivoting columns for analysis, and creating calculated columns such as crop categories (Cereals, Fruits, Beverages).
Data Type Conversion	Data types were rectified by converting production quantity columns to numeric format and year columns to whole numbers. This ensured accurate aggregation and calculations during analysis.

Column Splitting and Merging	Columns were split where crop names and categories were combined, and merged where entity-related information needed consolidation. This improved data readability and usability.
Data Modeling	Relationships were defined between tables based on common fields such as Year and Entity. DAX measures were created to calculate total production for crops like rice, wheat, maize, fruits, tea, and coffee. This enabled efficient aggregation and performance optimization.
Save Processed Data	The cleaned and transformed dataset was saved within the Power BI model and exported as a processed CSV file to ensure reuse for future analysis and reporting.