Data Collection and Preprocessing Phase

Date	06 October 2025
Team ID	SWUID20250214632
Project Title	Global Malnutrition Trends: A Power BI Analysis (1983-2019)
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 used for this project was sourced from Kaggle. It consists of two CSV files: 1. malnutrition-estimates.csv – Contains malnutrition indicators (stunting, wasting, underweight, overweight) for children under five years across multiple countries from 1983–2019. 2. country-wise-average.csv – Provides aggregated country-level averages for malnutrition indicators. The dataset includes details by country, region, and income classification (low, lower-middle, upper-middle, high income, and categories such as LDC, LIFD, LLDC, SIDS).
Data Cleaning	Checked for missing values in malnutrition indicators (e.g., stunting, underweight). 1. Removed irrelevant rows with incomplete or invalid entries. 2. Dropped duplicate country-year entries (if present). 3. Ensured consistency in country naming conventions.
Data Transformation	Using Power Query in Power BI: 1. Filtered the dataset to focus on children under five years of age. 2. Sorted data by year (1983–2019) for chronological analysis. 3. Created calculated columns such as:

	Total Underweight cases Total Overweight cases Percentage distribution by income level 4. Pivoted tables to allow year-wise and income-level comparisons.
Data Type Conversion	1.Converted Year column to Date/Integer type for proper chronological sorting. 2. Converted Survey Sample values to numeric type (integer). 3. Standardized malnutrition percentage values as decimal/percentage format.
Column Splitting and Merging	1.Split combined columns (if any country + year appeared together). 2.Merged relevant columns such as Country with Income Classification to simplify filtering in dashboard visuals.
Data Modeling	1.Established relationships between: □ country-wise-average.csv (Country → Average indicators) □ malnutrition-estimates.csv (Year → Indicators by Country) 2. Defined measures in Power BI for key metrics: □ Count of U5 Population □ Sum of Survey Samples □ Sum of Underweight Cases □ Average of Stunting by Income Level □ Overweight vs. Underweight comparison by Income Classification
Save Processed Data	▪ Saved the cleaned and transformed dataset in Power BI Data Model. ▪ Final .pbix file contains all pre-processing steps, calculated measures, and relationships for easy reproducibility.