

Data Collection and Preprocessing Phase

Date	01 OCT 2025
Team ID	XXXXXX
Project Title	Malnutrition Analysis and Forecasting 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 “malnutrition-estimates.csv” from Kaggle contains country-wise estimates of malnutrition indicators such as stunting, wasting, and underweight for various years. It includes columns like Country, ISO3 code, Year, Indicator, Value, and optionally Region. The dataset provides insights into malnutrition trends across countries, age groups, and genders.
Data Cleaning	Missing values in the Value column are handled by either removing rows with insufficient data or using interpolation for continuous years. Duplicate rows based on Country, Year, and Indicator are removed. Errors such as inconsistent country names are corrected using ISO3 mapping. Outliers or clearly incorrect values are flagged for review.
Data Transformation	Power Query in Power BI is used to filter relevant indicators, sort data by year and country, pivot/unpivot columns as needed for visualization, and create calculated columns such as Malnutrition Rate Change (current year – previous year). Measures for average malnutrition per country or region are also created.
Data Type Conversion	Columns are rectified to appropriate data types: Year → Whole Number, Value → Decimal Number, Country → Text, ISO3 → Text. Calculated columns and measures are automatically set to numeric types for analysis.

Column Splitting and Merging	If necessary, combined columns (e.g., “Indicator - Age Group”) are split for separate analysis. Country and ISO3 columns are merged where needed for mapping purposes in Power BI
Data Modeling	Relationships between tables (if multiple indicator tables exist) are defined using Country and Year as keys. Measures are created for total malnutrition, average value, and year-on-year change. Hierarchies are defined for Region → Country → Year for drill-down analysis.
Save Processed Data	The cleaned and processed dataset is saved as “malnutrition_cleaned.pbix” (Power BI file) and optionally exported as CSV for backup and future use. All transformations and measures are stored within the Power BI file.