Global Malnutrition Trends: A Power BI Analysis (1983-2019)

3.3. Data Exploration and Prepossessing

Date	19-10-25
Project Name	Global Malnutrition Trends: A Power BI Analysis (1983-2019)

1. Data Exploration

a. Dataset Overview

- Global Malnutrition Estimates (1983–2019) – Containing yearly malnutrition indicators such as stunting, underweight, and overweight prevalence among children under five.

- Country-Wise Averages Dataset – Summarized averages grouped by income classifications and regions.

b. Preliminary Analysis

- Assessed data dimensions (rows, columns, and field types).

- Identified missing or incomplete data in earlier years and less-reported countries.

- Examined distribution patterns of key indicators across countries and years.

- Detected outliers and anomalies (e.g., extremely high values in specific regions).

c. Initial Findings

- Stunting rates were generally higher in low-income countries.

- Overweight prevalence showed a gradual increase in middle- and high-income countries after 2000.

- Temporal trends revealed slow but steady improvement in underweight rates globally.

- A visible nutrition transition was observed — from under nutrition dominance to growing over nutrition issues.

Global Malnutrition Trends: A Power BI Analysis (1983-2019)

2. Data Preprocessing

a. Data Cleaning

- Removed duplicate entries.

- Handled missing values through interpolation or regional averaging.

- Standardized country names, region codes, and income classifications.

- Reformatted columns for uniformity (numeric, text, and date fields).

b. Data Transformation

- Converted percentage strings to numeric values for accurate calculations.

- Renamed columns for readability (e.g., "U5_Underweight" → "Underweight (%)").

- Created derived variables like:

  - Average Malnutrition Index

  - Total Malnutrition Burden by Region

  - Yearly Change (%)

c. Feature Selection

- Selected relevant attributes:

  - Country

  - Year

  - Income Group

  - Region

  - Indicators (Stunting, Underweight, Overweight)

- Removed redundant or non-essential columns that did not contribute to visualization.

d. Data Integration

- Merged both datasets using Country and Year as common keys.

- Established relationships for multi-dimensional analysis (e.g., country ↔ income group ↔ malnutrition rate).

3. Tools and Techniques Used

- Microsoft Excel: Used for initial exploration and descriptive statistics.

- Power BI Power Query: Applied for prepossessing , merging, and cleaning data.

- DAX (Data Analysis Expressions): Used for calculated columns and custom measures.

4. Data Validation

After prepossessing, validation checks were performed to ensure:

- All country-year combinations were consistent.

- No duplicate records existed.

- Summaries and totals matched source data set values.

- Trends and patterns aligned with known global nutrition statistics.

5. Outcome

After the exploration and prepossessing stage:

- Data was thoroughly examined, structured, and validated for accuracy.

- Clean and consistent datasets were created, ready for visualization in Power BI.

- Clear preliminary insights were drawn regarding global, regional, and income-based malnutrition patterns.

- This phase laid the groundwork for the next step — Data Visualization and Dashboard Design.