Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 14 June 2024 |
| 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** | Two datasets are used:   1. **country-wise-average.csv** — Provides average malnutrition indicators (Overweight, Underweight, Stunting) for each country. 2. **malnutrition-estimates.csv** — Contains yearly malnutrition   estimates (Underweight, Overweight, Stunting, Wasting) by country and year, categorized by sex and age group. |
| **Data Cleaning** | Missing values were found in key columns (e.g., Overweight, Underweight, Stunting). Duplicates in the **malnutrition- estimates.csv** were identified and removed based on Country- Year-Sex grouping. Country name inconsistencies across both  datasets were also resolved for proper joining. |
| **Data Transformation** | Power Query was used for filtering irrelevant rows, sorting data by country and year, pivoting malnutrition measures for summarization, and creating calculated columns such as "BMI  Gap" or "Malnutrition Score" where applicable. |
| **Data Type Conversion** | Data types were rectified:   * **Numeric:** Underweight, Overweight, Stunting, Wasting * **Text:** Country, Sex, Age group |

|  |  |
| --- | --- |
|  | - **Date/Number:** Year converted from Text to Integer for time  series analysis. |
| **Column Splitting and Merging** | Not required as columns were structured properly. However, where necessary, composite keys (like Country-Year) were  created for relationship building in the Data Model. |
| **Data Modeling** | Relationships between both datasets were defined using the **Country** column as the key. Calculated measures such as "Sum of Underweight," "Sum of Overweight," and "Average  Stunting by Year" were created using DAX formulas in Power BI. |
| **Save Processed Data** | The cleaned, transformed, and modeled datasets were saved  within Power BI Desktop for report generation and visualization in the Global Malnutrition Trends (1983–2019) dashboard. |