Group 12

Grourp Members

Ratul Saha	2020-1-60-036
Mahmuda Islam Rodela	2020-2-60-038
Al Mamun	2020-1-60-048

Description of Dataset(including question 1, 2,4,5)

Child Malnutrition - UNICEF Dataset

Global Expanded database Stunting, Wasting, Overweight

Dataset Description: This dataset is an approximate measure of Stunting, Wasting and Overweight. These are the symbols of Undernutrition. This dataset has 632(along with the title row) rows and 16 columns.

ISO: Three-digit alphabetical codes International Standard (ISO 3166-1) assigned by the International Organization for Standardization (ISO). It is representing the country codes of each country.

Countries and Areas: It is presenting the countries. There are 631 Not Null values are available in this column.

UNICEF Regions: This column representing country wise regions and its initials. It has no Null values.

Survey Years: Represents the years in which the Data collection took place.

Years: Year assigned to each survey for more information, consult the notes on the Data and Methodology in the key findings reports.

Short Source: It is needed for major survey types.

Full Source: Full title of the data source report.

Latest Estimate: This column represents the estimation of each country. The type of this column is categorical.

Estimate Type: This column's type of estimate if reanalyzed.

Wasted Only: Percentage of children under 5 years of age who are wasted only. A child with low weight for height is thin. Extreme thinness is called wasting. Number of children under 5 falling below -2 standard deviations from the median weight-for-height and falling at or above -2 standard deviations from the median height for age were counted.

Wasted and Stunted: Low weight for height is wasting and Low height for age is stunting. Number of children under 5 falling below -2 standard deviations from the median weight-for-height and falling below -2 standard deviations from the median height for age were counted for this column.

Stunted: This column representing the children's who has low height from their age. Number of children under 5 falling below -2 standard deviations from the median height for age and falling at or above -2 standard deviations from the median weight-for-height of the population.

Stunted and Overweight: Children's who are both stunted and overweighted. Number of children under 5 falling below -2 standard deviations from the median height for age and falling at or above +2 standard deviations from the median weight-for-height of the population.

Overweight Only: This column represents of children's who are overweight from their age and height.

Free from Wasting, Overweight, Stunting: This column is representing the percentage of children under 5 years of age who neither overweight, nor stunted, nor wasted. Number of children under 5 falling between -2 and +2 standard deviations from the median weight-for-height and falling at or above -2 standard deviations from the median height-for-age.

JME Country Estimates

Dataset Description: The Joint Malnutrition Estimates (JME) is a dataset of country estimates which is filled with the collection of national data. That contain information on child malnutrition. In this dataset, there are 1035 rows and 36 columns.

ISO Code: The ISO codes are internationally recognized standard codes which designated every country. It is like an initial that stands for a country or a state.

Country and Areas: It is presenting the countries. There are 1035 Not Null values available in this column.

United Nations Region: This column representing the names of the regions of those countries whose names have already written in this dataset.

United Nations Sub-Region: This column representing the names of the Sub-regions of those countries regions whose names have already written in this dataset.

SDG Region: Every region's Sustainable Development Goal (SDG) regions names are available in this column.

UNICEF Region: This column representing the regional initials of UNICEF Regions.

UNICEF Sub-Region: This column representing the Sub-regional initials of UNICEF Sub-Regions.

WHO region: World Health Organization's regions are designated in this column.

World Bank Income Classification: Bank Income classifications are here. It is an ordinal data type. There are 21 Null values available here.

World bank Region: This column representing the world bank regions.

LDC: Least development Countries (LDC) are low-income countries. This whole column representing least developed countries who has least income. It has 654 Null values that means these are developed than LDC.

LIFD: Low Income Food Deficient column indicating those countries who has low-income rating and food deficiency. There are 410 values available here that mean thee 413 countries has LIFD.

LLDC: Landlocked developing countries (LLDC) are developing countries that are landlocked. That countries don't have territory connected to an ocean. There are 376 landlocked countries available in this column.

UNICEF Survey ID: This column representing Survey IDs of UNICEF. The column is full of Not Null values.

WHO Global Database Number: This is presenting the database number.

Type of Estimate: There are different types of estimates are featured in this section.

WHZ Survey Sample (N): Weight for height survey samples is arranged in this column.

Severe Wasting: It is also known as Severe Acute Malnutrition (SAM). It is caused by lack of nutritious foods, frequent illness, poverty etc. It has 826 Not Null values.

Wasting: A child is low weight for height is thin and extreme thinness is called wasting. There are 990 Not Null values available in this column.

Overweight: Overweight refers to body weight that is greater than what is considered normal or healthy for a certain height.

Overweight Footnotes: This column representing some of the footnotes.

HAZ Survey Sample (N): Low height for are survey data's available here.

Stunting: Stunting is low height for age. There are 35 Null values are in this column.

Stunting Footnotes: This column mentioning some of the footnotes those footnote letters are available in Notes section. There are 446 values available here.

WAZ Survey Sample (N): Low Weight for age survey sample data placed in this column. There are 948 values available here.

Underweight: It is a combination of short- and long-term deficiency. Children's who have low weight for height and low height for age both are under weighted children.

Fieldwork Period: In this column, all the date, months, years are noted. There are 134 Null values which means they didn't do any fieldwork on those dates.

Report author: This column refers who published the dataset. There are no Null values here.

Source: This column refers where the datasets are collected from. The type of this column is nominal.

Short Source: This column refers the short forms of source institution.

U5 Population ('000s): Children's under 5 years of age are the members of Under 5 population. There are 1030 Not Null values available here.

Malnutrition Across The Globe

Country wise average

Dataset Description: In this dataset there are 8 (eight) attributes and 152 (one hundred fifty-two) rows. The attributes are Country, Income Classification, Severe Wasting, Wasting, Overweight, Stunting, Underweight, U5 Population.

Country: There are 152 unique values which represent the country name. That is the primary key of this dataset. The value of this column is nominal.

Income Classification: There are 4 values which represents Low Income: 0, Lower Middle Income: 1, Upper Middle Income: 2, High Income: 3. The value of this column is discrete.

Severe Wasting: Severe wasting, also known as severe acute malnutrition, is its most deadly form. It is caused by a lack of nutritious food and repeated bouts of diseases such as diarrhea, measles and malaria, which compromise a child's immunity. The value of this column is continuous.

Wasting: Wasting refers to who is too thin for his or her height and is the result of recent rapid weight loss or the failure to gain weight. The value of this column is continuous.

Overweight: Overweight refers to body weight that is greater than what is considered normal or healthy for a certain height. The value is of this column is continuous.

Stunting: Stunting is the impaired growth and development that children experience from poor nutrition, repeated infection, and inadequate psychosocial stimulation. The value of this column is continuous.

Underweight: Underweight is defined as low weight-for-age. The value of this column is continuous.

U5 Population: U5 Population refers population of children under 5 years old in a country. The value of this column is continuous.

Malnutrition Estimates

Dataset Description: In this dataset there are 20 (twenty) attributes and 924 (nine hundred twenty four) rows. The attributes are: Serial Number,

ISO code, Country, Survey year, Year, Income Classification, LDC, LIFD, LLDC or SID2, Survey Sample, Severe Wasting, Wasting, Overweight, Stunting, Underweight, Notes, Report Author, Source, Short source, U5 Population.

Serial Number: The value of this column is discrete. It starts from 0. This is the primary key of this dataset.

ISO code: The ISO country codes are internationally recognized codes that designate every country and most of the dependent areas a two-letter combination or a three-letter combination; it is like an acronym that stands for a country or a state. The value of this column is nominal.

Country: Country column represents the country name of the ISO code. The value of this column is nominal.

Survey year: This column refers which year the survey was done. The value of this column is discrete.

Year: This column refers which year the dataset was published. The value of this column is discrete.

Income Classification: There are 4 values which represents Low Income: 0, Lower Middle Income: 1, Upper Middle Income: 2, High Income: 3. The value of this column is discrete.

LDC: Least developed countries (LDC) are low-income countries confronting severe structural impediments to sustainable development. There are 3 values which represents Low: 0, Lower Middle: 1, Upper Middle Income: 2. The value of this column is discrete.

LIFD: In Load-Independent Flow Division column there are 3 values which represents Low: 0, Lower Middle: 1, Upper Middle Income: 2. The value of this column is discrete.

LLDC or SID2: The landlocked developing countries (LLDC) are developing countries that are landlocked. There are 3 values which represents Low: 0, Lower Middle: 1, Upper Middle Income: 2. The value of this column is discrete.

Survey Sample: This column refers how many sample was collected to estimate the survey. The value of this column is discrete.

Severe Wasting: Severe wasting, also known as severe acute malnutrition, is its most deadly form. It is caused by a lack of nutritious food and repeated bouts of diseases such as diarrhea, measles and malaria, which compromise a child's immunity. The value of this column is continuous.

Wasting: Wasting refers to who is too thin for his or her height and is the result of recent rapid weight loss or the failure to gain weight. The value of this column is continuous.

Overweight: Overweight refers to body weight that is greater than what is considered normal or healthy for a certain height. The value is of this column is continuous.

Stunting: Stunting is the impaired growth and development that children experience from poor nutrition, repeated infection, and inadequate psychosocial stimulation. The value of this column is continuous.

Underweight: Underweight is defined as low weight-for-age. The value of this column is continuous.

**Notes:** This column refers how the estimation was done. The value of this column is nominal.

Report Author: This column refers who published the data set. The value of this column is nominal.

Source: This column refers where the dataset was collected from. The value of this column is nominal.

Short source: This column refers the short form of the source institution. The value of this column is nominal.

U5 Population: U5 Population refers population of children under 5 years old in a country. The value of this column is continuous.

Pictorial Tree of our Dataset
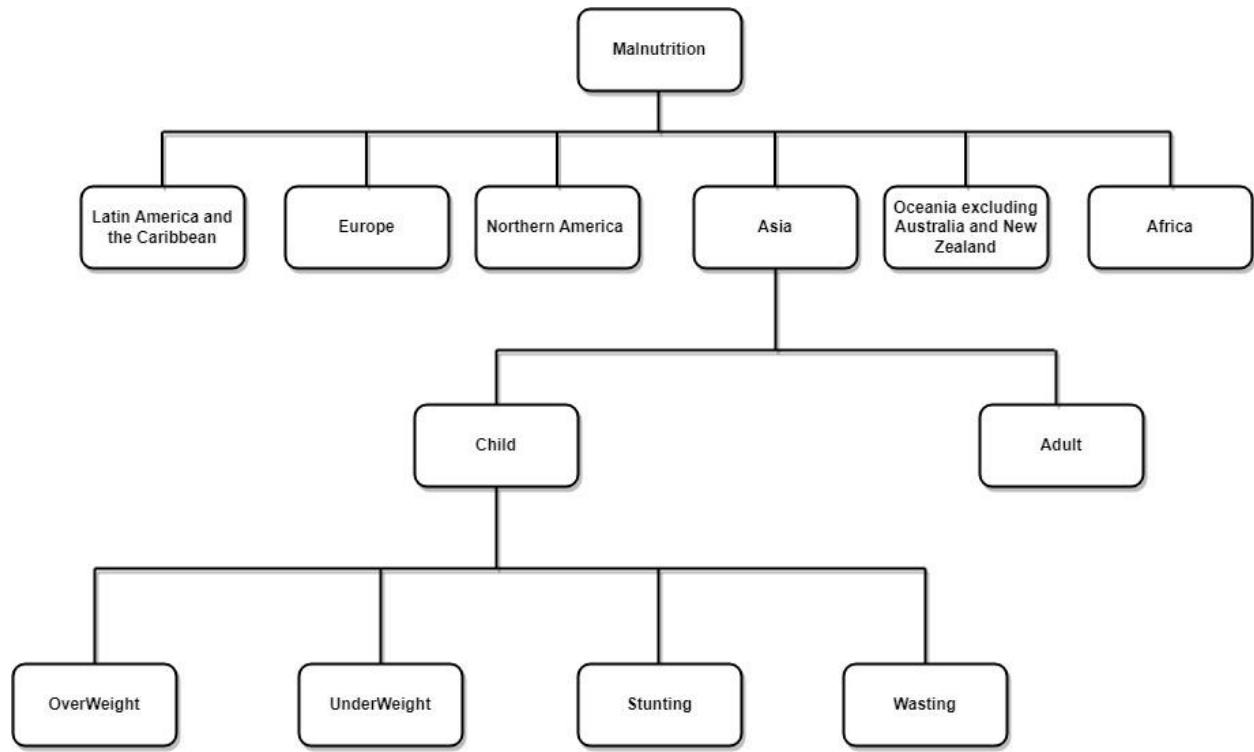
Malnutrition Dataset

1. Unicef Dataset for child: https://data.unicef.org/topic/nutrition/malnutrition/

2. Malnutrition across the globe: https://www.kaggle.com/datasets/ruchi798/malnutrition-across-the-globe

Research peper that has Citated These Datasets

1. https://www.sciencedirect.com/science/article/abs/pii/S0305750X0500077X

2. https://apps.who.int/iris/bitstream/handle/10665/341135/9789240025257-eng.pdf?sequence=1

3. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1574-0862.2001.tb00063.x

Worked analysis with these existing Dataset(Based on popularity)

1. https://www.kaggle.com/code/pauljef/data-visualization-child-malnutrition

2. https://www.kaggle.com/code/chanakyavivekkapoor/eda-using-bokeh-unicef-dataset

3. https://www.kaggle.com/code/vineethakkinapalli/unicef-child-malnutrition-eda/report

4. https://www.kaggle.com/code/tanyajain3108/how-can-we-prevent-malnutrition-in-children

5. https://www.kaggle.com/code/neelambarik/nutrition-analysis-of-countries

6. https://www.kaggle.com/code/priyanka1singh/country-wise-average-basic-eda-data-visualization

7. https://www.kaggle.com/code/mohitkr05/malnutrition-data-visualization

8. https://www.kaggle.com/code/pratik1120/malnutrition-the-disease-that-no-one-cares-about

9. https://www.kaggle.com/code/lily1917/correlation-exploration

Survey On Malnutrition Dataset

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9194519.4/

2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8971868/

3. https://peerj.com/articles/380/

4. https://onlinelibrary.wiley.com/doi/full/10.1111/mcn.13060

5. https://www.researchgate.net/profile/Palanichamy-Sudha/publication/329012021_Identification_Of_Malnutrition_With_Use_Of_Supervised_Data mining_Techniques-Decision_Trees_And_Artificial_Neural_Networks/links/5bef967b299bf1124fd82cec/Identificati on-Of-Malnutrition-With-Use-Of-Supervised-Datamining-Techniques-Decision-Trees-And-Artificial-Neural-Networks.pdf

6. https://aspenjournals.onlinelibrary.wiley.com/doi/abs/10.1002/jpen.1499

Drive Link of all Resources:
https://drive.google.com/drive/folders/1uvbFrlH0NnCDe7a8YQNj44GWwjdBJo4N?usp=sharing