**Diagram

Description automatically generated**

**Data Portfolio- Data Engenering**

By Maksat Mametjumayev

The article "Hunt and Undernourishment," which was posted on "Our World in Data," website served as the foundation for the database that was developed and the database that was analyzed. The article was written by Max Roser and Hannah Ritchie The database is primarily based on the second article's second section's final subcategory- "undernourishment." I focused on the data from the graphs in the "Depth of food deficit" category.

The primary measure of hunger according to FAO metrics is the prevalence of undernourishment. Although it shows us how many people fall below the minimal energy requirements, this measurement does not indicate how severely undernourished they are on average. The FAO used a statistic known as the "depth of the food deficit" to estimate the severity of undernourishment in a population. This measurement gives an estimate of the number of calories a typical person would require to maintain a healthy balance between their caloric intake and energy needs The average daily calorie deficit per person is expressed in kilocalories. My first table is based on the collected data about the “Depth of the food deficit in kilocalories per person per day, 2016”.

My second table was based on the data of the “Number of People Who are Undernourishment” which illustrates the total number of people who meet the criteria for being undernourished worldwide. Over the past few decades, there has been a declining, though erratic, the tendency on a global scale. The overall number of malnourished people has been declining. But during the past few years, the overall number climbed, reaching over 663 million in 2017.

Lastly, my three tables are associated with three crucial physiological indicators of child malnutrition and undernutrition. The following subsections will examine and illustrate the following measures:

* Stunting – being ‘too short for one’agee’. To be specific, the World Health Organization's Child Growth Standards define stunted children as having a height that is two standard deviations below the median height for age.
* Wasting – being ‘dangerously thin for one’s height’. In other words, wasting is a condition in which a person is dangerously underweight for their height and is typically an indication of fast weight loss, especially in youngsters. If a kid's weight for height is more than two standard deviations below the median for the international reference population of children aged 0-59 months, the child is considered wasted.
* Underweight – low weight-for-age in children. Undernourishment, or the prevalence of being underweight for age, might include children who are stunted, wasted or experiencing inadequate energy intake over a prolonged period of time.

So in short, I created a database that consists of 5 tables:

1. Food deficit
2. Number of undernourishment
3. Prevalence of stunning
4. Prevalence of underweight
5. Prevalence of wasting

I took several steps to evaluate the undernourishment data:

* The first step was *altering all tables to assign their primary keys* to each of the tables. The primary key was assigned as not null and auto increment.
* After that, I used the *select* command to get entity, year, and food deficit kilocalories per person per day from the food deficit table
* After using the select command I decide to get data about my country, Turkmenistan from the food deficit table. So I used *where* clause to specify my country name.
* Then I used *and, or, not* operators to compare Turkmenistan with Uzbekistan from the food deficit table.
  + - “And” operator gave empty rows- which means does not help for comparison
    - “Or” operator- returned all the data that relater returned all the data related to Turkmenistan and Uzbekistan which means it was an effective approach for the comparison
    - “Not” operator- returned all the data for all countries except Turkmenistan and Uzbekistan. So, not operator does not fit in our case.
* As the “or” operator gave needed output, I use it with other functions- order by and limit. When I ran the “or” operator, I got 50 rows. But I decided to sort the date from the newest to oldest and only get 20 rows. So, I order the year by descending and limited it by 20.
* Furthermore, I used “in and between” operators. I selected all the data from the prevalence of stunning table to get only data for the year- 2014,2015,2016. I choose all the information from the prevalence of stunning table between 2010 and 2018 for the "between" operator and ordered year by descending.
* I created three tables by building aggregate functions and using the inner join operation for all the tables:

1. Min\_value. I selected the entity and year from the food deficit table and conduct the “min” function for food deficit kilocalories per person per day, number of undernourishment(number of undernourishment table), height for age(prevalence of stunning), weight for age(prevalence of underweight), weight for height (prevalence of waste)
2. Max\_value. I chose the entity and year from the food deficit table and conduct the “max” function for food deficit kilocalories per person per day, number of undernourishment, height for age, weight for age, and weight for height.
3. Avg\_value. I picked the entity and year from the food deficit table and conduct the “avg” function for food deficit kilocalories per person per day, number of undernourishment, height for age, weight for age, and weight for height.

* Moreover, I implemented the union all command to combine the result of the two select commands and used “distinct” to delete duplicates as the union all generates duplicates.
* To get all records from food deficit and a number of undernourishment, I used the cross join operator.
* I also used a subquery to return data that will be used in the main query as a condition to further restrict the data to be retrieved. For instance, I selected entity, height for age, and weight for age in the main query as the average weight for age (inner query) selected only less or equal to the average weight for age which refers to the condition of the outer join.
* In addition, I implemented the case command in my analysis. I wanted to comment on the food deficit and the emergency of taking action. So, I claimed when the food deficit kilocalories per person per day is low or equal to 241.56, it is urgent which means needs to take action as soon as possible. When food deficit kilocalories per person per day is high or equal to 241.56, it is not urgent which means no needs to take action as soon as possible. However, if the output does not match these conditions, then there will be a comment “no ignorance” which means do not ignore the issue.
* Lastly, I used “locate and position” to get all countries that end with the “stan” suffix or countries whose suffix starts after 5th letter.

**References:**

Roser, M., & Ritchie, H. (2019). *Hunger and Undernourishment*. Our World in Data. <https://ourworldindata.org/hunger-and-undernourishment>