Individual Capstone Project
Deliverable #3: Report & Technical Analysis
By: Tina Trimmis
April 11, 2024

Topic:

Analyzing (the assumed rises) in price for healthy food categories, the affordability of a healthy diet and food insecurity around the World and some of its effects on those countries.

Approach Overview: Proposed Methods & Techniques:

1. Frame (Problem Statement/Guiding Question):

- The problem is the rising cost of food, particularly the affordability of a healthy diet.
- As food prices historically rise, there is increasing concern about access to nutritious foods, which is important for lowering the risk of diet-related illnesses and keeping the general public healthy.
- The objective of this data analysis capstone project is to examine the trends in healthy food category prices, assess the affordability of a healthy diet for different countries, and to hopefully advocate for policy recommendations to combat food insecurity and overall encourage a healthier diet amongst the population.
- A success metric would be to see governments/countries/organizations to implement policies to help this issue and to eventually see a decrease in food insecurity over time.

2. Extract (Data Collection):

Where is my data coming from?
I began by gathering data from reliable sources such as:

- The World Bank
    - An international financial institution that provides loans/grants to the governments of low/middle-income countries for the purpose of pursuing capital projects
    - Food Prices for Nutrition: https://databank.worldbank.org/source/food-prices-for-nutrition
    - *"Food Prices for Nutrition provides indicators on the cost and affordability of healthy diets in each country, showing the population's physical and economic access to sufficient quantities of locally available items for an active and healthy life. It also provides indicators on the cost and affordability of an energy-sufficient diet and a nutrient-adequate diet. These indicators are explained in detail in the Food Prices for Nutrition DataHub here: https://www.worldbank.org/foodpricesfornutrition."* - quote from their site
    - At the time of my download, this database had been last updated on 12/15/2023.
    - Some limitations to this data is that it only has the years 2017-2021, and the dataset was small and the bulk of the data is for 2017.

-

- Food and Agriculture Organization of the United Nations - The Food and Agriculture Organization Corporate Statistical Database (FAOSTAT)
    - FAO, a specialized agency of the United Nations that leads international efforts to defeat hunger
    - Suite of Food Security Indicators dataset: https://www.fao.org/faostat/en/#data/FS
    - Cost & Affordability of Healthy Diet (CoAHD) dataset: https://www.fao.org/faostat/en/#data/CAHD
    - At the time of my download, this database had been last updated in 2023 I believe.
- These datasets include data and statistics on the cost of a healthy diet, healthy diet affordability, and food insecurities all across various countries
- After exploring the Our World in Data datasets, I noticed they held the same information as FAO's Food Prices for Nutrition dataset, so I decided to eliminate this one since it was redundant.
- There is more specific information on these datasets mentioned in the third and fourth step of my data analysis process.

3. Prepare (Data Cleaning/Transformation):

- Clean the collected data to fix any nulls, errors, or outliers, renaming of columns, fix data types, make new columns if necessary, check for duplicates, and make it overall usable for analysis.
- I conducted 90% of my data cleaning in the attached Jupyter Notebooks in Python and about 10% in Tableau by creating calculated fields, dimensions, measures and parameters.
- The below outlines the steps I took in Python, which you can follow along in my Jupyter Notebooks for more detail.

Data Cleaning - Food Price for Nutrition (World Bank):

The Food Price for Nutrition dataset by The World Bank was a much smaller dataset than the FAO one, originally (3720, 40). I began my data transformation process by dropping 2 columns, 'Country Code 'Time Code'. Then I renamed the remaining column titles to lowercase and replace blank spaces in the names with _. I checked and converted data types appropriately. In Tableau I created calculated fields to filter the country column by region, income and only countries. Then I started tackling the nulls. I used numpy to replace values that had '..' for nulls, so they could be recognized as nulls..

Due to the set up and layout of this dataset, and since there was so much more data for 2017, I made the logical choice to split the dataset into two. I named these dataframes 'first' with 2017 data only and second' with 2018-2021 data.

Then I revisited handling the remaining nulls for each dataframe. For the 2017 "first" dataset, I dropped the empty rows that have 124 nulls in the below columns, because that meant they were all empty:

'percent_population_cannot_afford_sufficient_calories',
'percent_population_cannot_afford_nutrient_adequacy',
'percent_population_cannot_afford_healthy_diet',
'millions_people_cannot_afford_sufficient_calories',
'millions_people_cannot_afford_nutrient_adequacy',
'Millions_people_cannot_afford_healthy_diet'

And since these columns hold crucial data, they were not worth holding on to even though parallel columns held values because with nulls in these specific columns, made it not useful to hold onto for the purpose of this project. Then I replaced 24 rows of NaN with the word 'Unknown'. Because I didn't want to completely remove the rows since there were still some columns that held crucial values.

For the “second” dataframe with 2018-2021 data, handling the nulls was more straightforward. I dropped the columns for the second dataset (2018-2021) where all columns are completely empty (columns with 2472 null values). Then for the remaining nulls, 'cost_of_a_healthy_diet_[cohd]', 8 null rows for Argentina and Zimbabwe 2018-2021 (in FPN 2.0) have no data except in the Population column, so I dropped those too. I also dropped the two column's with 336 null rows because its not giving us any data, without these two critical column's data (see below), we only would have population size and the typical cost of a healthy diet and affordability of a healthy diet in this country:
'percent_of_the_population_who_cannot_afford_a_healthy_diet_[cohd_headcount]'
'millions_of_people_who_cannot_afford_a_healthy_diet_[cohd_unafford_n]'

Then there were 72 nulls remaining in the ‘affordability_healthy_diet_ratio_cost_to_food_poverty_line’ column, and since World Bank had CPI data available, I chose to Impute for the 72 nulls via a Python function calculating the affordability of a healthy diet ratio for the 'affordability_healthy_diet_ratio_cost_to_food_poverty_line' column for the Ratio of 'cost_healthy_diet' column to Food Poverty Line per year. The data dictionary tells us that the ratio of the cost of a healthy diet to the $1.12 food poverty line (52% of the international poverty line of $2.15/day in 2017 PPP$). (Data available for 2017, 2018, 2019, 2020 and 2021.)

Post data cleaning, I have 620 rows and 38 columns for “first” FPN World Bank dataset and 2128 rows and 9 columns for “second” FPN World Bank dataset.

Lastly, saved them to csv:

```
first.to_csv('./final_datasets/wb_food_prices_for_nutrition_cleaned_2017.csv index=False)
second.to_csv('./final_datasets/wb_food_prices_for_nutrition_cleaned_2018_2021.csv index=False)
```

Data Cleaning - Food Security and Cost/Affordability of a Healthy Diet (FAO):

From the FAO website, I decided to narrow my scope to these two datasets, that I end up merging:

1. Suite of Food Security Indicators dataset:
   fao_food_security = pd.read_csv('./faostat/Food_Security_Data/Food_Security_Data.csv’)
2. Cost & Affordability of Healthy Diet (CoAHD) dataset:
   fao_coahd = pd.read_csv('./faostat/Cost_Affordability_Healthy_Diet_CoAHD/Cost_Affordability_Healthy_Diet_CoAHD.csv')

Merged them both into **merged_fao** dataset/dataframe:

```
merged_fao.to_csv('./final_datasets/merged_fao_food_security_and_cost_affordability_healthy_diet_coahd.csv', index=False).
```

Upon merging, I had 237,895 rows and 13 columns. Something to note, is that during my cleaning phase, I found out that some of the columns from FAO’s Cost & Affordability of Healthy Diet

(CoAHD) dataset, matched with some of the columns from World Bank's Food Prices for Nutrition dataset (see below):

Cost of a healthy diet (PPP dollar per person per day)
Cost of starchy staples
Cost of animal source foods
Cost of legumes, nuts and seeds
Cost of vegetables
Cost of fruits
Cost of oils and fats
Percentage of the population unable to afford a healthy diet (percent)
Number of people unable to afford a healthy diet (million)

My data cleaning process for the columns consisted of dropping 'Area Code (M49)' and 'Year Code' columns and renamed the 'Flag' column to 'flag_code'. Then renamed the remaining column titles to lowercase and replace blank spaces in the names with _. For easier user understanding, I duplicated the Flag column and wrote out a description in the new column. I had to make a new column called 'value_condition' in order to convert the 'value' column data types that had a < symbol in it. Therefore the final remaining columns are: 'area_code', 'area', 'item_code', 'item', 'element_code', 'element', 'year', 'unit', 'value', 'flag_code', 'note', 'flag', 'value_condition'.

For cleaning rows, I eliminated all of the rows in the "item" column containing data on 3 year averages because I just wanted to focus on annual values for the nature of this project and timeline we had, and also eliminated gross domestic product per capita. The data on 3-year averages that I deleted are listed here:

'Average dietary energy supply adequacy (percent) (3-year average)
Share of dietary energy supply derived from cereals, roots and tubers (kcal/cap/day) (3-year average)
Average protein supply (g/cap/day) (3-year average)
Average supply of protein of animal origin (g/cap/day) (3-year average)
Gross domestic product per capita, PPP, (constant 2017 international $)
Prevalence of undernourishment (percent) (3-year average)
Number of people undernourished (million) (3-year average)
Prevalence of severe food insecurity in the total population (percent) (3-year average)
Prevalence of severe food insecurity in the male adult population (percent) (3-year average)
Prevalence of severe food insecurity in the female adult population (percent) (3-year average)
Prevalence of moderate or severe food insecurity in the total population (percent) (3-year average)
Prevalence of moderate or severe food insecurity in the male adult population (percent) (3-year average)
Prevalence of moderate or severe food insecurity in the female adult population (percent) (3-year average)
Number of severely food insecure people (million) (3-year average)
Number of severely food insecure male adults (million) (3-year average)
Number of severely food insecure female adults (million) (3-year average)
Number of moderately or severely food insecure people (million) (3-year average)
Number of moderately or severely food insecure male adults (million) (3-year average)
Number of moderately or severely food insecure female adults (million) (3-year average)

Cereal import dependency ratio (percent) (3-year average)
Percent of arable land equipped for irrigation (percent) (3-year average)
Value of food imports in total merchandise exports (percent) (3-year average)
Average fat supply (g/cap/day) (3-year average)

Lastly, I checked and converted data types appropriately. In Tableau I created a calculated field to filter the 'area' column with regions and income and only countries.

After data cleaning, removing nulls, changing data types, dropping columns and rows, checking for duplicates, creating new columns, checking for outliers using matplotlib and seaborn, and renaming columns, I have 104,711 rows and 12 columns to work with.

4. Analyze (Exploratory Data Analysis/Statistical Analysis):

- What is the data telling us? Any patterns, trends, and insights?
- Exploratory Data Analysis (EDA) and statistical analysis to find patterns present in the data
- Trends in food prices and affordability
- Apply statistical techniques to analyze the relationships between healthy food category costs and healthy food diet cost information
- Create basic visualizations in python, such as histograms, bar charts, scatter plots, etc. to visualize the data insights

The World Bank Food Prices For Nutrition dataset has four classifications: 1.0, 1.1, 2.0 and 2.1.
For both World Bank datasets, I am focusing my analysis only on 'Food Price for Nutrition 2.1 classification', because it holds the latest data as of October 2023.

EDA - "first" dataframe - Food Prices for Nutrition (World Bank 2017 only):

I am measuring the values in the following columns:

affordability_healthy_diet_ratio_cost_to_food_expenditures
affordability_healthy_diet_ratio_cost_to_food_poverty_line
affordability_nutrient_adequate_diet_ratio_cost_to_food_expenditures
affordability_nutrient_adequate_diet_ratio_cost_to_food_poverty_line
affordability_energy_sufficient_diet_ratio_cost_to_food_expenditures
affordability_energy_sufficient_diet_ratio_cost_food_poverty_line
cost_healthy_diet
cost_healthy_diet_relative_to_cost_of_sufficient_energy_from_starchy_staples
cost_nutrient_adequate_diet 'cost_energy_sufficient_diet
cost_animal-source_foods
cost_animal-sourced_foods_relative_to_starchy_staples_in_least-cost_healthy_diet
cost_fruits
cost_fruits_relative_to_starchy_staples_in_least-cost_healthy_diet
cost_legumes_nuts_seeds
cost_legumes_nuts_seeds_relative_to_starchy_staples_in_least-cost_healthy_diet
cost_oils_fats
cost_oils_fats_relative_to_starchy_staples_in_least-cost_healthy_diet

cost_starchy_staples
cost_vegetables
cost_vegetables_relative_to_starchy_staples_in_least-cost_healthy_diet
cost_share_animal-sourced_foods_in_least-cost_healthy_diet
cost_share_fruits_in_least-cost_healthy_diet
cost_share_legumes_nuts_seeds_in_least-cost_healthy_diet
cost_share_oils_fats_in_least-cost_healthy_diet
cost_share_vegetables_in_least-cost_healthy_diet
cost_share_starchy_staples_in_least-cost_healthy_diet
percent_population_cannot_afford_sufficient_calories
percent_population_cannot_afford_nutrient_adequacy
percent_population_cannot_afford_healthy_diet
millions_people_cannot_afford_sufficient_calories
millions_people_cannot_afford_nutrient_adequacy
millions_people_cannot_afford_healthy_diet 'population

Top 5 countries with highest cost of healthy diet:

- Jamaica $5.98, Japan $ 5.53, Suriname $4.97, Korea $4.71, Guyana $4.63
- Lowest = UK $1.82

Millions of people who can't afford a healthy diet:

- India 1066.8, China 232.4, Indonesia 192.5, Pakistan 175.3, Nigeria 174.6

% of pop who cannot afford a healthy diet:

- Madagascar 97.1%, Burundi 95.8%, African republic 95.7%, Malawi 94.5%, Congo 94.2%
- On the lower end at 0.10% is Cyprus Finland France Slovenia United Arab Emirates
- The 3 with 0% Switzerland, Iceland, Azerbaijan.
- Countries with over 85.8% of the population who couldn't afford a healthy diet were all African countries.

All the healthy food category costs below are measured by dollars per person, per day. Jamaica has the most expensive Oils, Fats, Fruits and Vegetables. Japan has the most expensive starchy staples and second most expensive vegetables.

Cost of Fruits:

- Jamaica $2, Mongolia $1.92, Korea Rep. $1.33, Guyana $1.28, Tunisia $1.13
- Least: Belize $0.16, Zimbabwe $0.20, Sao Tome and Principe $0.24, Azerbaijan $0.26, Sierra Leone $0.30

Cost Legumes, Nuts, Seeds:

- Seychelles $0.88, South Africa $0.85, Paraguay $0.83, Suriname $0.78, Brazil $0 .74
- Least: Guinea-Bissau $0.07, Niger $0.09, Senega $0.12, Burkina Faso $0.13, Zimbabwe $0.13

Cost Oils Fats:

- Sri Lanka $0.32, Egypt $0.26, Honduras $0.25, Jamaica $0.24, Sudan $0.21
- Least: Germany, Denmark, Netherlands, Spain, UK all $0.05

Cost Vegetables:

- Guinea $1.78, Japan $1.57, Jamaica $1.45, Costa Rica $1.39, Panama $1.3
- Least: Rwanda $0.3, Azerbaijan $0.32, Madagascar $0.32, Moldova $0.34, Kazakhstan $0.34

Cost of Starchy Staples:

- Japan $1.48, Liberia $1.27, Thailand $0.90, Taiwan, China $0.86, Angola $0.84
- Least: UK $0.15, Finland $0.2, Luxembourg $0.21, Switzerland $0.21, Zimbabwe $0.21

Cost animal source foods:

- Haiti $1.58, Suriname $1.41, Nepal $1.38, Jordan $1.29, Central african republic $1.28
- Least: Zimbabwe $0.37, Senegal $0.38, Australia $0.46, UK $0.49, Denmark $0.5

Healthy Food Category Costs by Income:

- All the costs are measured by dollars per person, per day here too.
- The most expensive food category by country's Income level was Vegetables, across all income levels, especially Upper Middle Income & High Income.
- Fruit was the overall second most expensive, also for Upper Middle & High Incomes
- Starchy Staples was most expensive for Low Income Countries
- Legumes, Nuts and Seeds were most expensive for Upper Middle Income.
- Oils & Fats were most expensive for High Income Countries
- Overall – Upper Middle Income countries pay the majority of the highest costs.

When evaluating healthy food category costs by region, I found that no specific region carried the burden of highest costs for all healthy food categories, it was pretty evenly distributed, no regions repeating.

Overall affordability of a healthy diet: ratio of cost to food poverty line:

- Jamaica $5.34, Japan $4.94, Suriname $4.44, Korea rep. $4.21, Guyana $4.13
- Least: UK $1.63, senegal $1.96, zimbabwe $1.96, iceland $1.98, australia $2.02

In 2017, these are the following amounts worldwide:

- $3.263 - cost of a healthy diet
- $2.364 - cost of a nutrient adequate diet
- $0.7889 - cost of a energy sufficient diet
- $0.877 Cost of animal source foods
- $0.659 Cost of fruits
- $0.340 Cost of legumes, nuts and seeds
- $0.126 Cost of oils and fast
- $0.486 Cost of starchy staples
- $0.768 - Cost of vegetables
- The % of the pop that cannot afford sufficient calories = 6.665%
- The % of the pop that cannot afford nutrient adequacy - 26.984%
- The % of the pop that cannot afford healthy diet = 37.203%
- The millions of people that cannot afford sufficient calories is 9.06 mill.
- The millions of people that cannot afford nutrient adequacy is 59.194 mill.
- The millions of people that cannot afford a healthy diet is 80.86 mill.

EDA - “second” dataframe - Food Prices for Nutrition (The World Bank 2018-2021):

For 2018-2021, we have less columns/information in our dataset to explore.
We are measuring these four column categories:

Affordability_healthy_diet_ratio_cost_to_food_poverty_line
Cost_healthy_diet
percent_population_cannot_afford_healthy_diet
Millions_people_cannot_afford_healthy_diet

East Asia & Pacific ranks #1 for the most expensive cost of a healthy diet, every year, ranging from $3.70-$4.21. Followed by Latin America & Caribbean for second most expensive all years, and third most expensive was South Asia all 5 years. North America had some fluctuations in 2017 to 2018 dropped $0.07, 2020 to 2021 dropped $0.05. Other than North America during those two years, all other regions increased every year from 2017-2021.

Top 5 highest cost of healthy diet:

- 2018: Jamaica $6.14, Japan $5.70, Suriname $5.31, Mongolia $4.66, Korea rep $4.90
- 2019: Jamaica $6.398, Japan $5.565, Suriname $5.337, Mongolia $4.9, Korea rep $4.83
- 2020: Jamaica $6.68, Japan $5.647, Suriname $5.739, Mongolia $5.115, Korea rep $5.11
- 2021: Jamaica $7.03, Japan $5.638, Suriname $6.09, Mongolia $5.676, Korea rep $5.34
  Per year worldwide cost:
  - 2018 - $3.325
  - 2019 - $3.386
  - 2020 - $3.4776
  - 2021 - $3.623

Top 5 highest countries with Millions of people who can’t afford a healthy diet:

- 2018: India 1002, indonesia 190, nigeria 181, pakistan 175, china 191
- 2019: India 970, indonesia 187, nigeria 187, pakistan 182, china 172
- 2020: India 1038, indonesia 191, nigeria 194, pakistan 189, china 189
- 2021: India 1013, indonesia 194, nigeria 200, pakistan 185, china 129
  Per year worldwide:
  - 2018 - 79.108
  - 2019 - 78.356
  - 2020 - 81.738
  - 2021 - 82.005

Top 5 highest countries with % of pop who cannot afford a healthy diet:

- 2018: Madagascar 97.3%, central african rep. 95.4%, malawi 94.9%, burundi 95%, mozambique 92.8%
- 2019: Madagascar 97.1%, central african rep. 95.4%, malawi 95.4%, burundi 95%, mozambique 93.6%
- 2020: Madagascar 97.8%, central african rep. 95.6%, malawi 95.7%, burundi 95.7%, mozambique 94.4%

- 2021: Madagascar 97.8%, central african rep. 95.8%, malawi 95.9%, burundi 95.9%, mozambique 94.8%
- Per year worldwide:
    - 2018 - 36.505%
    - 2019 - 36.06%
    - 2020 - 36.924%
    - 2021 - 36.62%

Top 5 highest countries with overall highest affordability of a healthy diet: ratio of cost to food poverty line:

- 2018: Jamaica 5.48, japan 5.09, suriname 4.74, mongolia 4.166, korea rep 4.375.
- 2019: Jamaica 5.71, japan 4.969, suriname 4.765, mongolia 4.375, korea rep 4.315.
- 2020: Jamaica 5.965, japan 5.04, suriname 5.124, mongolia 4.567, korea rep 4.563.
- 2021: Jamaica 6.279, japan 5.034, suriname 5.438, mongolia 5.068, korea rep 4.768.
- Per year worldwide:
    - 2018 - 3.054
    - 2019 - 3.114
    - 2020 - 3.195
    - 2021 - 3.234

Overall in 2017, across the globe, the cost of a healthy diet was $3.27 per day. 37.42% of the population and over 800 million people could not afford a healthy diet.

EDA - “merged_fao” dataframe - Food Security & Cost/Affordability of a Healthy Diet (FAOSTAT)

The items/columns I am analyzing in this dataset are within the “item” column, listed as the below categories per country, region, income level.

1. Political stability and absence of violence/terrorism (index)
2. Per capita food production variability (constant 2014-2016 thousand int$ per capita)
3. Per capita food supply variability (kcal/cap/day)
4. Percentage of population using safely managed drinking water services (Percent)
5. Percentage of population using at least basic drinking water services (percent)
6. Percentage of population using at least basic sanitation services (percent)
7. Percentage of children under 5 years affected by wasting (percent)
8. Number of children under 5 years affected by wasting (million)
9. Percentage of children under 5 years of age who are stunted (modelled estimates) (percent)
10. Number of children under 5 years of age who are stunted (modeled estimates) (million)
11. Percentage of children under 5 years of age who are overweight (modelled estimates) (percent)
12. Number of children under 5 years of age who are overweight (modeled estimates) (million)
13. Prevalence of obesity in the adult population (18 years and older)
14. Number of obese adults (18 years and older) (million)

15. Prevalence of anemia among women of reproductive age (15-49 years)
16. Number of women of reproductive age (15-49 years) affected by anemia (million)
17. Minimum dietary energy requirement (kcal/cap/day)
18. Average dietary energy requirement (kcal/cap/day)
19. Percentage of population using safely managed sanitation services (Percent)
20. Prevalence of low birthweight (percent)
21. Number of newborns with low birthweight (million)
22. Prevalence of undernourishment (percent) (annual value)
23. Number of people undernourished (million) (annual value)
24. Prevalence of severe food insecurity in the rural adult population (percent) (annual value)
25. Prevalence of severe food insecurity in the total population (percent) (annual value)
26. Prevalence of severe food insecurity in the town and semi-dense area adult population (percent) (annual value)
27. Prevalence of severe food insecurity in the urban adult population (percent) (annual value)
28. Prevalence of severe food insecurity in the male adult population (percent) (annual value)
29. Prevalence of severe food insecurity in the female adult population (percent) (annual value)
30. Prevalence of moderate or severe food insecurity in the rural adult population (percent) (annual value)
31. Prevalence of moderate or severe food insecurity in the total population (percent) (annual value)
32. Prevalence of moderate or severe food insecurity in the town and semi-dense area adult population (percent) (annual value)
33. Prevalence of moderate or severe food insecurity in the urban adult population (percent) (annual value)
34. Prevalence of moderate or severe food insecurity in the male adult population (percent) (annual value)
35. Prevalence of moderate or severe food insecurity in the female adult population (percent) (annual value)
36. Number of severely food insecure people (million) (annual value)
37. Number of severely food insecure male adults (million) (annual value)
38. Number of severely food insecure female adults (million) (annual value)
39. Number of moderately or severely food insecure people (million) (annual value)
40. Number of moderately or severely food insecure male adults (million) (annual value)
41. Number of moderately or severely food insecure female adults (million) (annual value)
42. Cost of a healthy diet (PPP dollar per person per day)
43. Cost of starchy staples 'Cost of animal source foods
44. Cost of legumes, nuts and seeds 'Cost of vegetables
45. Cost of fruits 'Cost of oils and fats
46. Percentage of the population unable to afford a healthy diet (percent)
47. Number of people unable to afford a healthy diet (million)

The average across all countries and all years, per item category listed above, is as follows:

item	value
Average dietary energy requirement (kcal/cap/day)	2345.391122
Cost of a healthy diet (PPP dollar per person per day)	3.456844
Cost of animal source foods	0.874000
Cost of fruits	0.666468
Cost of legumes, nuts and seeds	0.348009
Cost of oils and fats	0.130645
Cost of starchy staples	0.505036
Cost of vegetables	0.787482
Minimum dietary energy requirement (kcal/cap/day)	1818.281244
Number of children under 5 years affected by wasting (million)	2.090828
Number of children under 5 years of age who are overweight (modeled estimates) (million)	0.868655
Number of children under 5 years of age who are stunted (modeled estimates) (million)	4.735584
Number of moderately or severely food insecure female adults (million) (annual value)	104.534291
Number of moderately or severely food insecure male adults (million) (annual value)	96.045402
Number of moderately or severely food insecure people (million) (annual value)	293.709337
Number of newborns with low birthweight (million)	0.617334
Number of obese adults (18 years and older) (million)	9.235228
Number of people unable to afford a healthy diet (million)	135.291649
Number of people undernourished (million) (annual value)	118.554417
Number of severely food insecure female adults (million) (annual value)	38.693852
Number of severely food insecure male adults (million) (annual value)	34.465805
Number of severely food insecure people (million) (annual value)	108.857796
Number of women of reproductive age (15-49 years) affected by anemia (million)	11.650652
Per capita food production variability (constant 2014-2016 thousand int$ per capita)	17.082822
Per capita food supply variability (kcal/cap/day)	37.394041
Percentage of children under 5 years affected by wasting (percent)	5.966176
Percentage of children under 5 years of age who are overweight (modelled estimates) (percent)	7.015160
Percentage of children under 5 years of age who are stunted (modelled estimates) (percent)	21.592174

Percentage of population using at least basic drinking water services (percent)	84.925076
Percentage of population using at least basic sanitation services (percent)	72.060432
Percentage of population using safely managed drinking water services (Percent)	67.375488
Percentage of population using safely managed sanitation services (Percent)	50.753292
Percentage of the population unable to afford a healthy diet (percent)	36.629228
Political stability and absence of violence/terrorism (index)	-0.058142
Prevalence of anemia among women of reproductive age (15-49 years)	28.051068
Prevalence of low birthweight (percent)	11.298816
Prevalence of moderate or severe food insecurity in the female adult population (percent) (annual value)	30.386973
Prevalence of moderate or severe food insecurity in the male adult population (percent) (annual value)	28.155460
Prevalence of moderate or severe food insecurity in the rural adult population (percent) (annual value)	33.007407
Prevalence of moderate or severe food insecurity in the total population (percent) (annual value)	29.163772
Prevalence of moderate or severe food insecurity in the town and semi-dense area adult population (percent) (annual value)	31.724444
Prevalence of moderate or severe food insecurity in the urban adult population (percent) (annual value)	29.434091
Prevalence of obesity in the adult population (18 years and older)	16.573482
Prevalence of severe food insecurity in the female adult population (percent) (annual value)	10.861973
Prevalence of severe food insecurity in the male adult population (percent) (annual value)	9.958046
Prevalence of severe food insecurity in the rural adult population (percent) (annual value)	12.532593
Prevalence of severe food insecurity in the total population (percent) (annual value)	10.409897
Prevalence of severe food insecurity in the town and semi-dense area adult population (percent) (annual value)	11.763704
Prevalence of severe food insecurity in the urban adult population (percent) (annual value)	10.495455
Prevalence of undernourishment (percent) (annual value)	10.865909

Severe Food Insecurity data collected was from 2014 to 2022. The prevalence of Severe Food Insecurity by region annually, ranges: 1% to 39%. Middle Africa & Caribbean had it the worst from 2020-2022. USA, Europe, Oceania, and some regions of Asia have the least (all years). In more than half of the regions, severe food insecurity grows annually, instead of getting better.

For data on the prevalence of Low Birthweight, its latest data was recorded in 2020. Range was from 4% - 27.4%. India had the highest low birth weights, followed by Comoros 23%, Philippines 21.1%, Liberia 19.9%, Nepal 19.7%, Madagascar 18.7%. Iceland had the lowest low birth weights.

Population per country having at least basic versus safe drinking water:
The data provided a range for this from 2000 to 2020. I wanted to see if the percentages of the population with safely managed drinking water services and if those using at least basic drinking water services have improved. The % of the population with this basic need ranges from 2.3% to 99%.
2000: Uganda had the worst with 2.3% of population using safely managed.
2020: Chad had the worst with 5.6% of population using safely managed.
2000: Ethiopia had the worst with using at least basic drinking water services 18.1%.
2020: Ethiopia still had the worst with basic at 49.6%.

There are a handful of countries that had the highest percent, at 99% in 2000 and the same in 2020 such as the UK, USA, United Arab Emirates, etc for at least basic. And again, UK, Sweden, Spain and many others had 99% for safely managed in 2000 and 2020 for using safely managed

Population of at least basic versus safe sanitation services:
The data provided a range for this from 2000 to 2020. I wanted to see if the percentages of the population with safely managed sanitation services and those using at least basic sanitation services and the improvement in those countries. The percentages ranged from 2.10% to 99%. For at least basic sanitation services, in 2000 Ethiopia had 2.1% (the worst) and Madagascar 4.2%. For the best, it was the USA, UK, Switzerland and various European countries. In 2020, Ethiopia only improved to 8.9% Madagascar to 12.3%. The map shows us that South America and Mexico and Australia improved. For countries using safely managed sanitation services, in 2000, only 4 countries with 99%, Switzerland, Singapore, Kuwait, and Austria. USA has 97.5% In 2000, Ethiopia had 2.1% and Madagascar 4.4% and in 2020: Ethiopia 6.7%, Madagascar 10.4% (lowest countries). In 2020, the United Arab Emirates, Switzerland, Singapore all had 99% and the USA had 98.3%.

36.6% was the average percent of the population that could not afford a healthy diet in 2021.
In 4 years, the number has only dropped less than a percent.

Some Conclusions:

- The cost of healthy food groups in order from highest cost to lowest for ALL Countries was Fruits, Vegetables, Animal-Source, Starchy Staples, Legumes/nuts/seed, and least expensive was oils/fat.
- Upper Middle Income Countries paid the majority of the highest costs for healthy food categories in 2017.
- All regions except for North America, increased their cost of healthy foods every year 2017-2021.
- African countries came up a lot in my EDA.
- East Asia & Pacific Region had the highest cost of a healthy diet every year from 2017-2021.
- Jamaica had the most expensive cost of a healthy diet 2017-2021.
- In 2017, Jamaica had the Oils and Fats and Fruits and also the most expensive Vegetables.
- Japan had the second most expensive diet. Japan has the most expensive starchy staples and 2nd most expensive vegetables.

5. Interpret:

- Connect findings to the original problem statement and overarching analysis question

6. Communicate (Share insights & recommend a course of action):
    - Create visuals
    - Create a user friendly dashboard
    - Storytelling
    - Recommendations
    - Presentation with a slide deck consisting of insights, visuals, and recommendations for a non-technical audience (stakeholders, the general public, potential policymakers to implement changes)

Next Steps:

- Didn't get include all of my visuals/analysis in my presentation
- Clean up Jupyter Notebooks, make them neater.
- More EDA to do with my current datasets especially with the FAO dataset:
- Look more into Affordability of a Healthy Diet Ratio of Cost to Food Poverty Line, etc.
- Add Visual about Obesity, Undernourishment, moderate food security etc.
- Spruce up visuals in Tableau
- Improve Dashboards
- Create a Story in Tableau
- Predictive Model
- Come up with overall Recommendations
- Post on Medium/Tableau Public
- Data Clean/EDA on the other relevant datasets I collected, such as FAO's below, may try to merge them into the **merged_fao** dataset.:
    1. Consumer Price Indices
    2. Food Balance Sheets
    3. Food Prices