

Individual Capstone Project
Deliverable # 2: Topic Proposal and Dataset
By: Tina Trimmis
March 24, 2024

Topic:

Analyzing (the assumed rise of) Food Prices and the Affordability of a Healthy Diet

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 food 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 security over time.

2. Extract (Data Collection): Where is the data coming from?

- I gathered data from reliable sources such as:
 - Our World in Data
 - A non-profit organization based in the UK
 - The Food and Agriculture Organization
 - FAO, a specialized agency of the United Nations that leads international efforts to defeat hunger
 - 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
- These datasets include data and statistics on the cost of a healthy diet, healthy diet affordability, and food insecurities all across various countries
- There is more specific information on these datasets at the end of this document

3. Prepare (Data Cleaning/Transformation):

- Clean the collected data to fix any nulls, errors, or outliers, renaming of columns, fix data types, and make it overall usable for analysis.

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

- What is the data telling us?
- EDA and statistical analysis to find any patterns present in the data
- Trends in food prices and affordability

- Apply statistical techniques to analyze the relationships between healthy food costs and healthy diet information
 - Create basic visualizations in python, such as histograms, bar charts, scatter plots, etc to visualize the data insights
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
 - 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)

Data:

Using 3 Datasets (see below in bold for file names). I have more datasets from FAO, but not sure if I will have time to use them by the final deadline; may try to merge them into the merged FAO dataset.

1. Our World In Data:

- a. Cost of a healthy diet as a share of average expenditure on food, 2017:
<https://ourworldindata.org/grapher/cost-healthy-diet-share-food-expenditure>
- b. Daily cost of a healthy diet, 2017-2021:
<https://ourworldindata.org/grapher/cost-healthy-diet>
- c. Share of the population who cannot afford a healthy diet percent, 2017-2021:
<https://ourworldindata.org/grapher/share-healthy-diet-unaffordable>
- d. Number of people who cannot afford a healthy diet percent, 2017-2021:
<https://ourworldindata.org/grapher/number-healthy-diet-unaffordable>

* Merged the above four into one dataset **"owid_final.csv" (886 rows, 7 columns)**

2. The World Bank Group:

- a. Food Prices for Nutrition: <https://databank.worldbank.org/source/food-prices-for-nutrition>

* **"wb_Food_Prices_for_Nutrition.csv" (3,725 rows, 40 columns)**

3. Food and Agriculture Organization of the United Nations - The Food and Agriculture Organization Corporate Statistical Database (FAOSTAT):

- a. Suite of Food Security Indicators: <https://www.fao.org/faostat/en/#data/FS>
- b. Cost & Affordability of Healthy Diet (CoAHD): <https://www.fao.org/faostat/en/#data/CAHD>

* Merged two datasets: "Food_Security_Data.csv" + "Cost_Affordability_Healthy_Diet_CoAHD.csv" into **"merged_fao_food_security_and_coahd.csv" (237,895 rows, 13 columns)**