



Checkpoint I: Project Proposal

Group: G09

Date: 2022/10/04

Domain

Since 1975, worldwide obesity has nearly tripled. In 2016, more than 1.9 billion adults, 18 years and older, were overweight. Of these over 650 million were obese. Globally, one billion people, including 1 in 5 women and 1 in 7 men, are estimated to be living with obesity by 2030. The fundamental cause of obesity and overweight is an energy imbalance between calories consumed and calories expended.

The idea of this project is to give detailed information about various food items. The main objective is to make people aware about the quality of food that they usually consume and encourage them to choose a healthier option. We intend to study many characteristics of a large group of food items which provide information about whether the item is rich in protein, has too much salt or has a high calorie density, has a high percentage of salt, among others.

We hope that this visualization helps people put their eating habits into perspective and persuades them to improve their diet.

Dataset

Dataset description:

- Which dataset will we be using? We will be using a dataset found on [Kaggle](#) which contains specific nutrition values for around 8.8k types of food. We did some modifications, in particular the "name" column which was partitioned in two other columns (type and description) since it has a lot of relevant information that can be analyzed independently. We also added a column "category" that classifies the item between Fruits, Vegetables, Meat, Fish, Oils, Dairy, Starchy food, Fast food, Beverages and Others. Finally, we decided to add a table that has the recommended daily intakes of the calories, macronutrients, sodium, potassium and sugar for an adult.
- How will we verify the reliability and accuracy of the values? We decided to check some of the values present in this dataset to guarantee that they are reliable and accurate. After multiple iterations through various items, the results proved to be trustworthy.

Dataset:

<https://www.kaggle.com/datasets/trolukovich/nutritional-values-for-common-foods-and-products>

Example Questions

This dataset has a total of 77 columns, consequently we have numerous data entries where we can do many queries and understand detailed information about each food item as well as compare different ones. Since there are many attributes for each item, we can start asking questions like how attributes can be related to each other and how one attribute can influence the others.

Given the domain and our goal with this dataset, some adequate questions that we aim to answer are:

- Do food items with a high protein percentage and low calories per serving generally have a low percentage of saturated fat?

- How does the total fat and sodium influence the cholesterol of an item?
- Is fish generally less caloric than meat with a low fat percentage?
- Does fiber rich starchy food tend to have less cholesterol?

Soluble fiber forms a gel-like substance in your intestines, slowing down digestion. It also traps cholesterol and prevents your body from reabsorbing it into your bloodstream. Given that information, it is relevant to understand if food items with the most fiber are specially poor in cholesterol.

- Is water rich food healthier?

The concept of healthy food can be defined by a good balance of the various macronutrients: carbohydrates, fats and protein in comparison with the reference table.

- Does caffeine reduce the presence of B-vitamins, vitamin C and minerals in the food?

Caffeine has a mild diuretic effect, which leads to an increase in urination. As a result, water-soluble vitamins, such as B-vitamins and vitamin C can be depleted due to fluid loss. Researches also demonstrate that the higher the level of caffeine, the more it interferes with vitamin D and sodium absorption.

Based on these facts, with this question we want to try to establish a relationship between the amount of caffeine and the presence of vitamins in a food item itself.

Data Sample

This example of data can show that all the above questions are adequate and valid.

Ex: (from "nutrition.csv")

```
,name,type,description,category,serving_size,calories,total_fat,saturated_fat,cholesterol,sodium,choline,folate,folic_acid,niacin,pantothenic_acid,riboflavin,thiamin,vitamin_a,vitamin_a_rae,carotene_alpha,carotene_beta,cryptoxanthin_beta,lutein_zeaxanthin,lucopene,vitamin_b12,vitamin_b6,vitamin_c,vitamin_d,vitamin_e,tocopherol_alpha,vitamin_k,calcium,copper,iron,magnesium,manganese,phosphorous,potassium,selenium,zinc,protein,alanine,arginine,aspartic_acid,cystine,glutamic_acid,glycine,histidine,hydroxyproline,isoleucine,leucine,lysine,methionine,phenylalanine,proline,serine,threonine,tryptophan,tyrosine,valine,carbohydrate,fiber,sugars,fructose,galactose,glucose,lactose,maltose,sucrose,fat,saturated_fatty_acids,monounsaturated_fatty_acids,polyunsaturated_fatty_acids,fatty_acids_total_trans,alcohol,ash,caffeine,theobromine,water
3,"Teff, uncooked","Teff,uncooked,Vegetables",100 g,367,2.4g,0.4g,0,12.00 mg,13.1 mg,0,0,3.363 mg,0.942 mg,0.270 mg,0.390 mg,9.00 IU,0.00 mcg,0.00 mcg,5.00 mcg,0.00 mcg,66.00 mcg,0,0,0.482 mg,0,0,0.08 mg,0.08 mg,1.9 mcg,180.00 mg,0.810 mg,7.63 mg,184.00 mg,9.240 mg,429.00 mg,427.00 mg,4.4 mcg,3.63 mg,13.30 g,0.747 g,0.517 g,0.820 g,0.236 g,3.349 g,0.477 g,0.301 g,0,0.501 g,1.068 g,0.376 g,0.428 g,0.698 g,0.664 g,0.622 g,0.510 g,0.139 g,0.458 g,0.686 g,73.13 g,8.0 g,1.84 g,0.47 g,0.00 g,0.73 g,0.00 g,0.01 g,0.62 g,2.38 g,0.449 g,0.589 g,1.071 g,0,0,2.37 g,0,0,8.82 g
```

(from "reference.csv")

```
age,tot_calories,protein(g),protein(kcal),carbohydrates(g),carbohydrates(kcal),fat(g),tot_fat(kcal),sat_fat(kcal),sugar(kcal),potassium(g),sodium(g)
```

```
adult,2000,84.86,300,340,1300,62.9,600,200,200,0.3,5
```