



Checkpoint I: Project Proposal

Group: G09

Date: 2022/09/23

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. 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, has a high calorie density or is rich in macrominerals, for example.

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](https://www.kaggle.com/datasets/trolukovich/nutritional-values-for-common-foods-and-products) which contains specific nutrition values for around 8.8k types of food. We intend to do some modifications, in particular the "name" column which will be partitioned since it has a lot of relevant information.
- 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 high protein and low calories per serving generally have a low percentage of saturated fat?
- How does the total fat influence the cholesterol of an item?
- Are macrominerals and trace minerals abundant in low calorie food items?

- Does fiber rich 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.

- Does caffeine affect the presence of vitamins 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,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
2,"Eggplant, raw",100 g,25,0.2g,,0,2.00 mg,6.9 mg,22.00 mcg,0.00 mcg,0.649 mg,0.281 mg,0.037 mg,0.039 mg,23.00 IU,1.00 mcg,0.00 mcg,14.00 mcg,0.00 mcg,36.00 mcg,0,0.00 mcg,0.084 mg,2.2 mg,0.00 IU,0.30 mg,0.30 mg,3.5 mcg,9.00 mg,0.081 mg,0.23 mg,14.00 mg,0.232 mg,24.00 mg,229.00 mg,0.3 mcg,0.16 mg,0.98 g,0.051 g,0.057 g,0.164 g,0.006 g,0.186 g,0.041 g,0.023 g,0,0.045 g,0.064 g,0.047 g,0.011 g,0.043 g,0.043 g,0.042 g,0.037 g,0.009 g,0.027 g,0.053 g,5.88 g,3.0 g,3.53 g,1.54 g,0,1.58 g,0,0,0.26 g,0.18 g,0.034 g,0.016 g,0.076 g,0.00 mg,0.0 g,0.66 g,0.00 mg,0.00 mg,92.30 g
```