

SAE 2.04 – nutriscore

Timothe Jacob et Kenzo Ruffet-Ejiri

05/04/2024

Introduction

The objective of this project is to find out what are the key factors that affect the Nutriscore of a product, especially within the processed meats in the United States.

But first, what is Nutriscore?

The Nutri-Score is a labeling system designed to provide consumers with easily understandable information about the nutritional quality of a food product. It assigns a letter grade and color code ranging from A (green) to E (red) based on the overall nutritional composition of the product, considering factors such as energy, sugars, saturated fats, sodium, protein, fiber, and the presence of fruits, vegetables... This system allows consumers to make healthier food choices by quickly assessing the nutritional value of a product at a glance.

We have been given a dataset of many products and their information.

To answer the question, we will study the data of processed meats products, produced in the United States.

Data set Presentation :

First we cleaned and filtered the “openfoodfacts” database (link : <https://world.openfoodfacts.org/>), and we only kept the useful columns, therefore our data set contains the following attributes :

- code (product id),
- url (links to the product’s page),
- product_name (name of the product),
- brands_tags (the linked brand),
- stores (the linked stores),
- owner (the linked owner),
- food_groups (the type of food – **filtered by processed meat**),
- labels_tags (information about the allergens),
- countries (linked countries – filtered to keep only the **USA**),
- countries_tags (tags of the countries),
- quantity (the quantity of the product),
- fat_100g (the fat presence percentage),
- saturated_fat_100g (the saturated fat presence percentage),
- sugars_100g (the sugars presence percentage),
- proteins_100g (the proteins presence percentage),
- carbohydrates_100g (the carbohydrates presence percentage),
- energy_100g (the energy in kJ),
- salt_100g (the presence of salt in percentage),
- sodium_100g (the presence of sodium in percentage),
- nutriscore_score (the score of the nutriscore (in integer)),
- nutriscore_grade (the grade of the nutriscore (A,B,C,D,E)),

- `computed_energy_100g` (the energy calculated),
- `organic` (if it's organic or not),
- `vegan` (if it's vegan or not),
- `vegetarian` (if it's vegetarian or not),
- `palm_oil` (if it contains palm oil or not),
- `level_fat` (the level of fat (low, moderate, high))
- `level_saturated_fat` (the level of saturated fat (low, moderate, high))
- `level_sugars` (the level of sugars (low, moderate, high))
- `level_salt` (the level of salt (low, moderate, high))

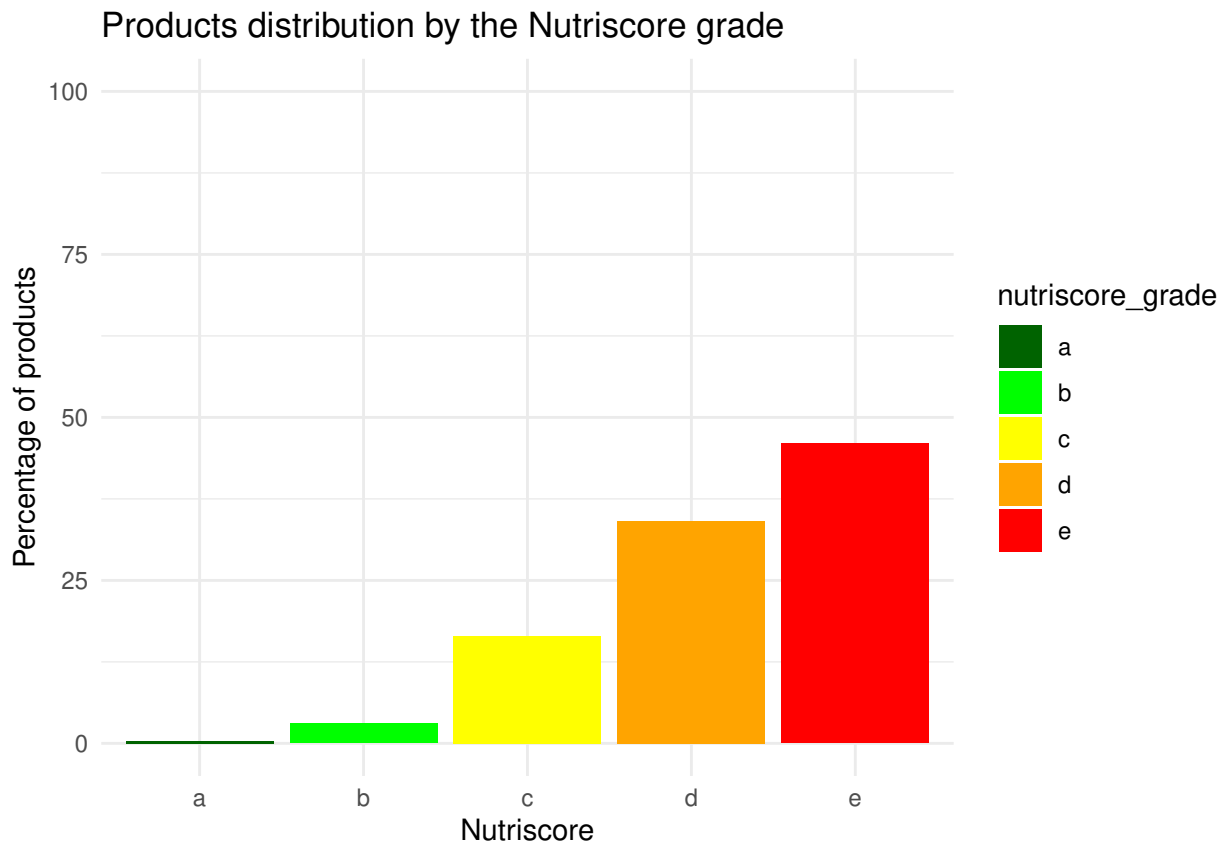
```
## [1] 6977
```

As we can see, there are 6977 entries in our data set. It's enough to be representative.

Analysis

Percentage of products for each score

First, let's take a look at the percentage of each Nutriscore



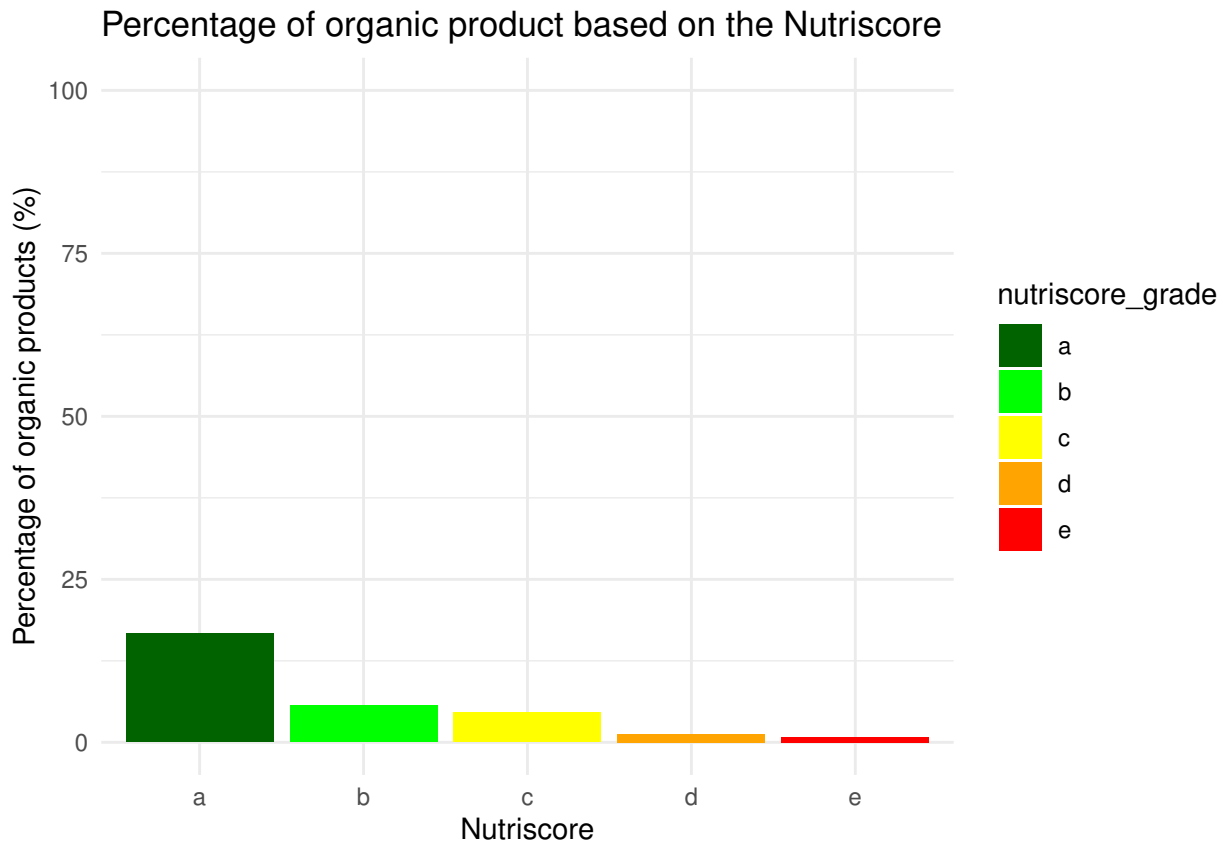
Here, we see that almost half of the products have a Nutriscore of E, which is the worst score. It is followed by the D score, with around 35% of the products. Then the C score, with 15% of the set, and finally, both the B and A scores are under 5%.

That sets the mood : majority of processed meats are badly rated. It's noticeable that each time the score gets worse, there's an increase in the quantity of products of that score.

But what makes them rated so bad? Or maybe, we should reverse the question : what makes an A-rated product, rated so good? Do organic products have better Nutriscores?

The share of organic products based on the Nutriscore

Here's a bar graph showing, in percentage, the share of product that are organic based on their Nutriscore :

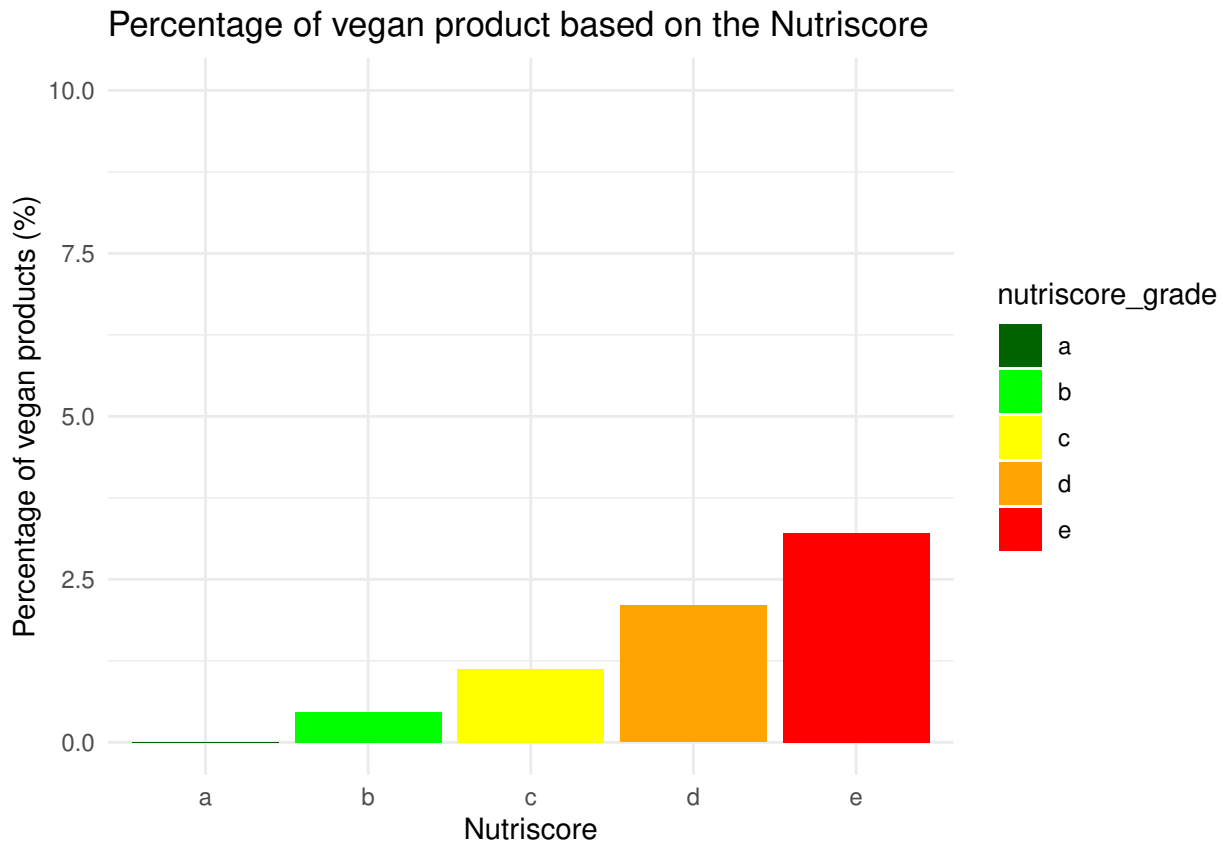


We can see here that, for the A-rated product, we have almost 20% of them that are organic. As we go to the worst scores, we see that the percentage of organic products is lowering again and again. So, we deduct that organic products tend to have a better Nutriscore than non-organic products.

But what about vegan products? Do they have a better Nutriscore?

Vegan products share based on the Nutriscore

There's a graph showing the share of products that are vegan for each Nutriscore.



These stats are radically different : we see that each time the Nutriscore gets worse, there are more vegan products. We scaled the y-axis to 10% due to the really low amount of vegan products :

`## Number of vegan products : 167 out of 6977`

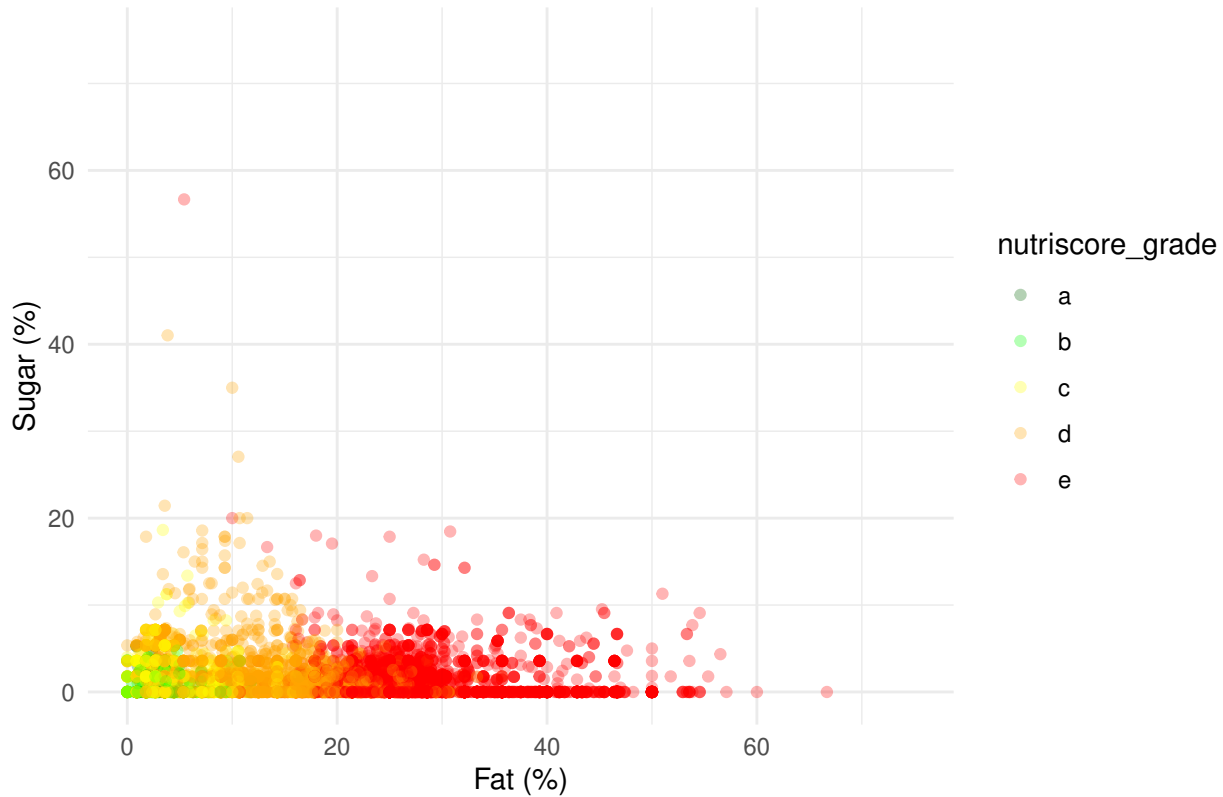
In fact, there are only 167 vegan products. And despite the low number, we see that they essentially have bad Nutriscores. And it seems pretty logical : we are doing this study on **meat** products. To make food that tastes like meat with no animals, it needs to be stuffed with tons of chemicals and other components that tends to lower the Nutriscore grade.

But maybe bad Nutriscores are also linked with fat or sugar... let's find out :

Fat and sugar influence on Nutriscore

Here is a dot-chart showing the Nutriscore of a product by its fat percentage and its sugar percentage. To read it, each dot represents a product, and is colorized by it's Nutriscore. The closer the product is to the top, the more sugar it contains. And the closer it is to the right, the more fat it contains.

Fat vs Sugar influence on Nutri-Score



The distribution of these products shows that Nutriscore is affected by its fat percentage. On the other hand, sugar percentage doesn't seem to affect it a lot, as most red dots are on the right.

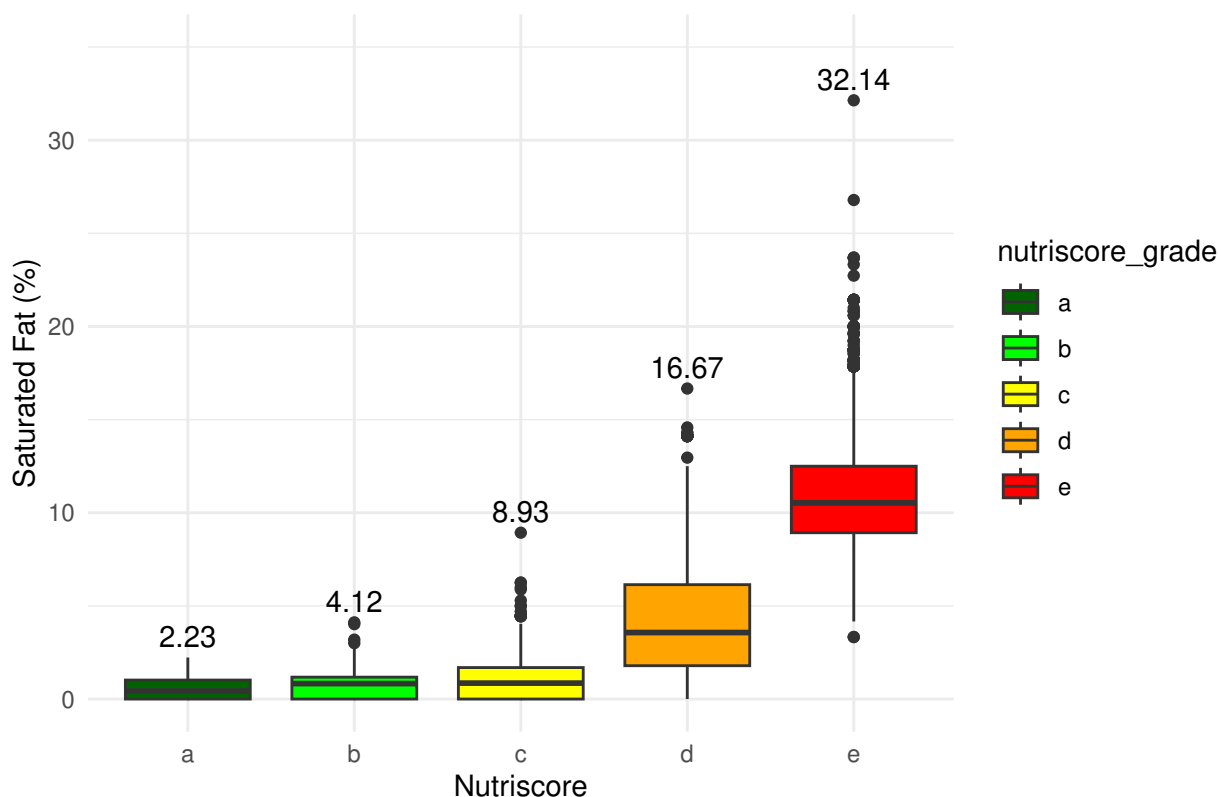
That means that **fat is a more influential factor than sugar**.

But what about saturated fat? Does it have a more important role in the attribution of the Nutriscore?

Saturated fat influence on the Nutriscore

There's a graph showing 5 boxplots (one per Nutriscore) showing the involvement of saturated fats in the notation.

Saturated fat percentage per Nutriscore



This is an interesting graph. First, by looking at the y-axis, we see that the limit is around 30%, which means that, out of all the products selected for this analysis, none of them exceeds this limit. So we know that the average percentage of saturated fat in these products is relatively low :

Average of saturated_fat_100g : 6.58056416090996

Only 6.5% of average presence of saturated fats. However, looking at the graph, we can see that they do indeed influence the Nutriscore, even if the values are apparently low. In fact, if we take a closer look at the A score, we note that there are no extreme values, and none of the products have a higher value than 2.5%. The B score seems to be justified by the extreme values, that go up to 4.12%. And it's repeating again, and again.

We can admit that **saturated fats** are also **influential** on the Nutriscore, even if the score seems in appearance pretty low.

Conclusion

Based on the analysis conducted, several key insights emerge regarding the Nutriscore of processed meat products in the United States.

Firstly, the majority of products in the dataset have low Nutriscores, with nearly half rated as E, indicating poorer general nutritional quality. This prevalence of low scores suggests that a significant portion of processed meat products may not align with nutritional guidelines.

Organic products tend to have better Nutriscores, with a higher proportion of A-rated products being organic compared to those with lower grades. This indicates a potential link between organic production methods and better nutritional quality.

Conversely, there is a notable increase in the proportion of vegan products as Nutriscores decrease. This trend suggests that while vegan options may appeal to consumers seeking healthier alternatives, processed vegan meat substitutes may not consistently achieve higher Nutriscores, possibly due to the use of additives or processing methods.

Finally, the difference analysis on fats and saturated fats reveals that they appear to have a more significant influence on Nutriscores compared to other, like sugar content. Products with higher fat percentages tend to receive lower Nutriscores, highlighting the importance of fat reduction in improving nutritional quality.