



## Report SAE2.04

2024-04-09

## Table of contents

<b>1. Introduction .....</b>	<b>3</b>
1.1 Percentage breakdown of products by Nutri-Score .....	3
1.2 Question answer.....	4
<b>2. Analyze of the data set.....</b>	<b>4</b>
2.1. Import of the data set .....	4
2.2. Descriptive analysis of the data set.....	4
Sugar content.....	5
Salt content.....	7
Saturated fat graph.....	8
Proteins graph.....	10
Energy graph .....	12
Boxplot Conclusion.....	13
Bio, vegetarian, vegan .....	14
Palm oil.....	16
<b>3. Conclusion.....</b>	<b>17</b>

Figure 1 :Percentage of products per Nutri-Score.....	3
Figure 2 : Sugar content distribution of products by score Nutri-Score .....	5
Figure 3: Average sugar content per Nutri-Score .....	6
Figure 4 : Salt content per Nutri-Score.....	7
Figure 5 : Saturated fat content distribution of products by Nutri-Score.....	8
Figure 6 : Proteins content distribution of products by score Nutri-Score.....	10
Figure 7 : Average protein content per Nutri-Score .....	11
Figure 8 : Average energy content per Nutri-Score .....	12
Figure 9 : Comparison of sugar, proteins, saturated fat and energy content of products by Nutri-Score .....	13
Figure 10 : Percentage of vegan products.....	14
Figure 11 : Percentage of organic products.....	14
Figure 12 : Percentage of products containing palm oil by Nutri-Score.....	16

## 1. Introduction

The nutriscore is a nutrition label that classifies food products according to their nutritional quality. The nutriscore is calculated based on the nutritional composition of the product. In our case it's base on sugar, saturated fat, salt, proteins, and energy. The nutriscore is a color-coded scale from A to E, with A being the best score and E the worst. The nutriscore is intended to help consumers make healthier food choices.

The type of product analyzed in this study is biscuits and cakes. The objective of this study is to analyze the nutritional quality of biscuits and cakes sold in the United States. The analysis is based on the nutriscore of the products. Before analyzing the data, we have already cleaned up the data by removing missing values and duplicates.

We will explore 22400 products of biscuits and cakes sold in the United States. The data set contains 13 variables, including the product name, the brand, the nutriscore, the energy content, the fat content, the saturated fat content, the sugar content, the salt content...

### 1.1 Percentage breakdown of products by Nutri-Score

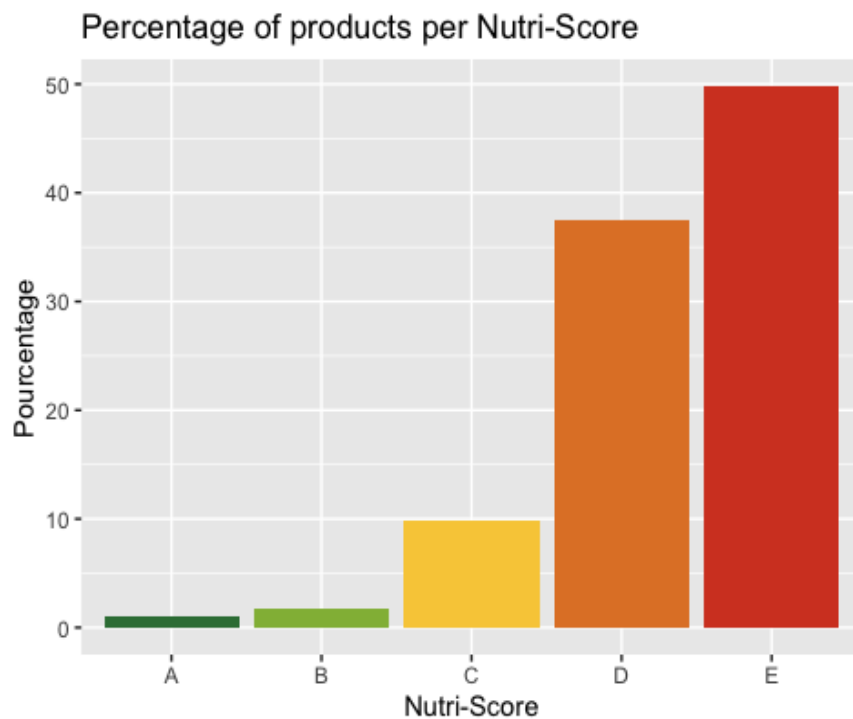


Figure 1 :Percentage of products per Nutri-Score

The bar chart above shows the percentage of products by Nutri-Score. On the x-axis, we have the Nutri-Score, and on the y-axis, we have the percentage of products.

Here, it is important to note that most products have a grade E nutriscore, around 50%. Next, around 37% of products have a grade D nutriscore. Approximately 10% of products have a grade C nutriscore. Nearly 2% of products have a grade B nutriscore. Finally, less than 1% of products have a grade A nutriscore. Only three percent of products have a nutriscore A or B, worrying information.

Therefore, we can observe that most products in the biscuits and cakes category sold in the United States have a grade D or E nutriscore (87%), which means that the majority of products in this category have mediocre nutritional quality. This is both important and alarming data, which will be useful for the rest of our analysis.

## 1.2 Question answer

The aim of analyzing these data is to answer the following question : What link can be established between the nutriscore and nutritional variables for products in the '*en:biscuits-and-cakes*' family in the United States ?

## 2. Analyze of the data set

### 2.1. Import of the data set

To answer this question, we are going to analyze the distribution of nutritional variables according to the nutriscore. In this way, we will analyze the distribution of products according to their sugar, salt, fat and energy content. We will then compare the distribution of products according to whether they are organic, vegetarian, vegan or contain palm oil.

To begin with, we imported the dataset. The dataset is stored in a csv file named "clavagua.csv".

### 2.2. Descriptive analysis of the data set

Here we will look at an overall analysis of the products observed according to the Nutri-Score. We will look at the sugar, salt, fat and energy content of products for each Nutri-Score.

Before starting our observations, it's important to notice that :

- a product is Nutriscore A if its score is less than -1,
- a product is Nutriscore B if its score is between 0 and 2,
- a product is Nutriscore C, if its score is between 3 and 10,
- a product is Nutriscore D, if its score is between 11 and 18,
- a Nutriscore E product if its score is above 19.

### Sugar content

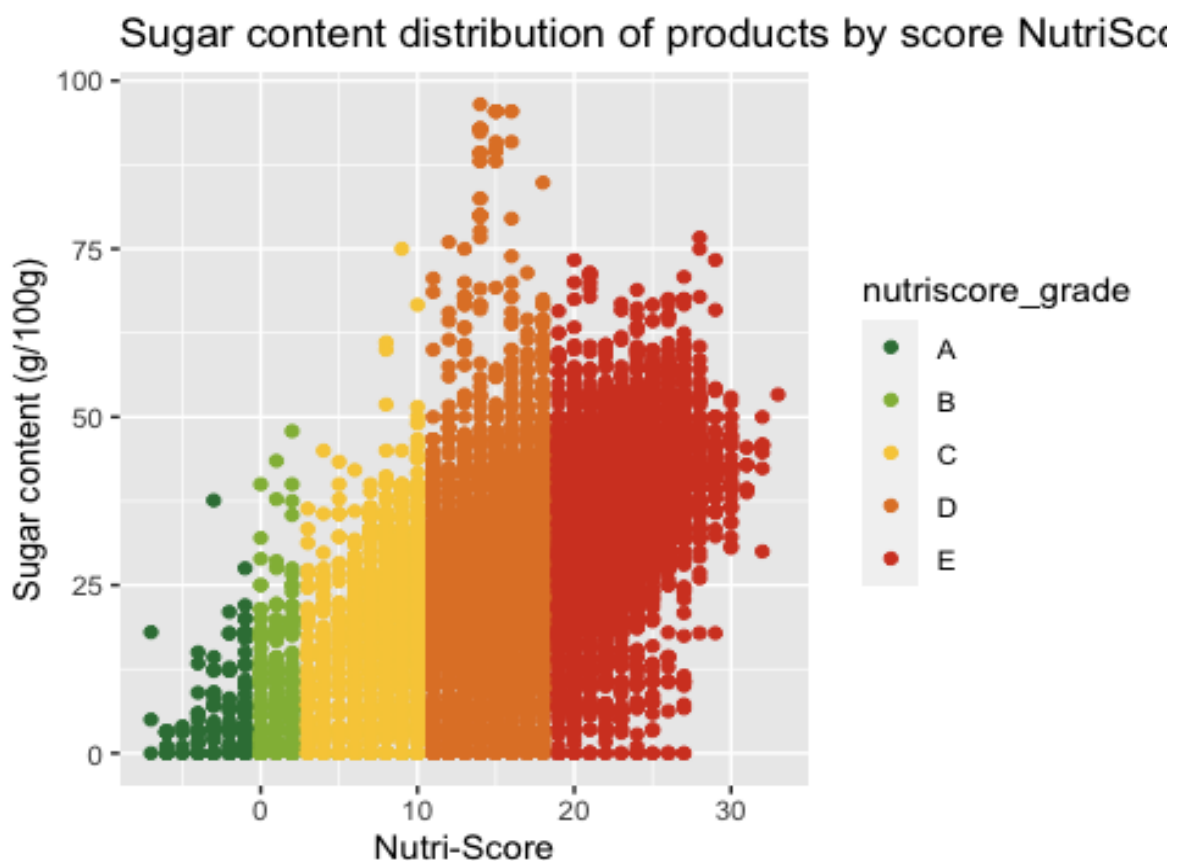
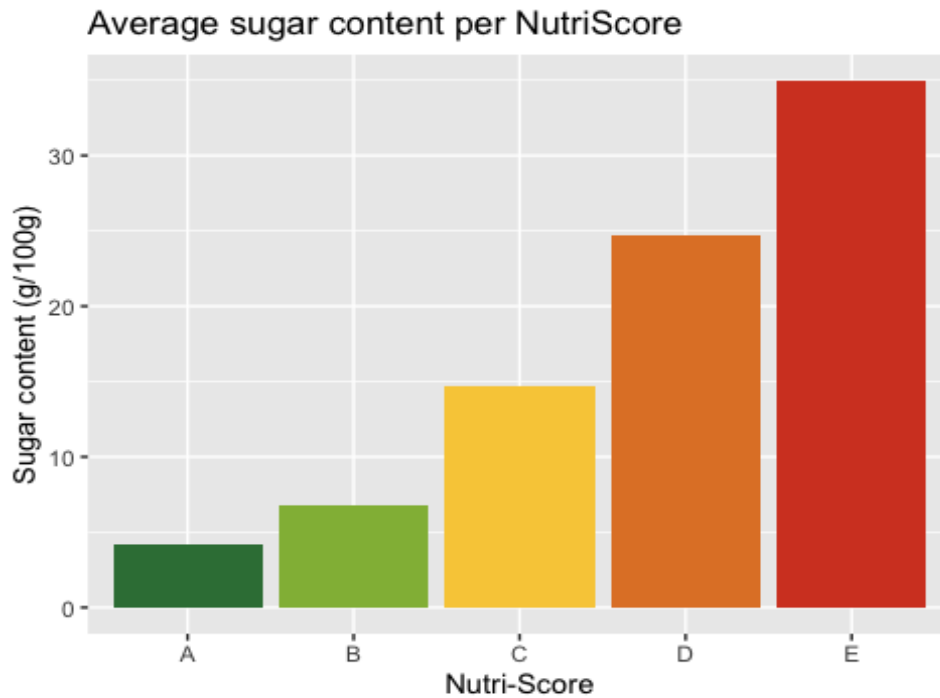


Figure 2 : Sugar content distribution of products by score Nutri-Score

The first graph is a scatter plot showing the sugar content of products according to the Nutri-Score. On the x-axis, we have the score for each product according to its nutritional characteristics. On the y-axis, we have the specific content of sugar. Finally, we have colored the points according to the Nutri-Score. The first thing we can see is a general upward trend in the sugar content of products, depending on the score. The higher score the product, the

higher the sugar content. We can see that products with a healthy nutriscore tend to have a lower sugar content. We can also see, for example, that the product with the highest nutriscore A has a sugar content of around 37.5g/100g. In contrast, the product with the highest nutriscore E has a sugar content of around 75g/100g, i.e. twice as much. This shows a clear difference between healthy and less healthy products in terms of sugar content.

We can also notice that the type of product with the highest sugar content (around 90g/100g) is generally Merengue. This is a type of cake made from egg whites and sugar, which explains the high sugar content.



*Figure 3: Average sugar content per Nutri-Score*

The second graph is a bar chart showing the average sugar content per Nutri-Score. On the x-axis, we have each Nutri-Score and on the y-axis, we have the mean of sugar content for each Nutri-Score. It confirms what we observed before. Products with a high nutriscore tend to have a higher average sugar content. Thus, nutriscore E products reach peak sugar content of around 35g/100g, more than double the sugar content of nutriscore C products, five times the sugar content of nutriscore B products and seven times the sugar content of nutriscore A products.

##	nutriscore_grade	high	medium	low
##	A	2	26	209
##	B	18	85	274
##	C	614	634	931
##	D	5281	1638	1318
##	E	9831	814	300

We just created a table with the number of products with a high, medium, and low sugar content for each Nutri-Score. Most products have a high sugar content, regardless of the Nutri-Score. This is worrying information, as high sugar content is bad for health. For example, if we take the nutriscore E, we can see that the part of “high” sugar content represents 32 times the part of “low” sugar content. This is a very high difference, which shows that the majority of products in this category have a high sugar content. However, if we take the nutriscore A, we can see that the part of “low” sugar content represent 100 times the part of “high” sugar content. This means that the majority of products in this category have a low sugar content. The closer a product is to nutriscore A, the healthier its composition.

### Salt content

Before any analysis, it is important to note that we are analyzing cakes, biscuits, and sweet products, which have a logical tendency to have a low salt content. We do not attach great importance to the salt content of sweet products. However, we are going to analyze the salt content of products according to the Nutri-Score.

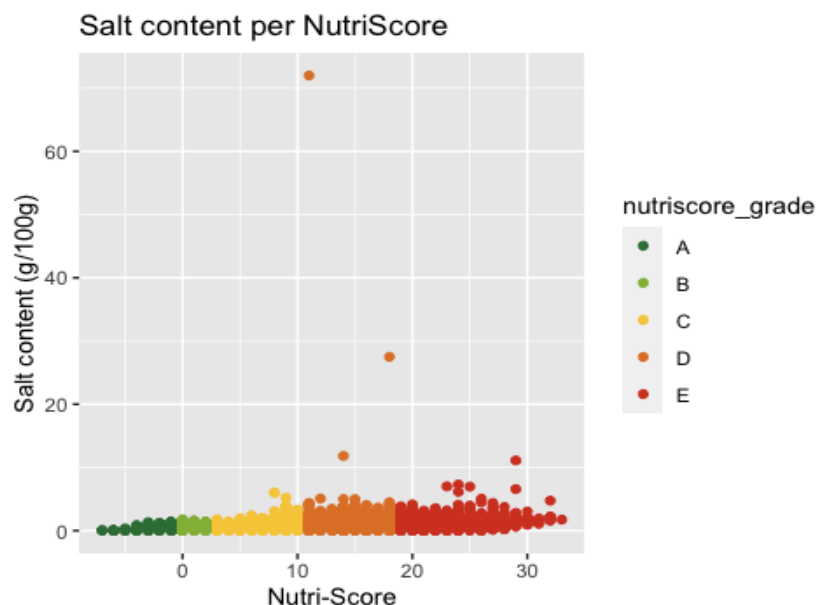


Figure 4 : Salt content per Nutri-Score

This graph is a scatter plot showing the salt content of products as a function of the Nutri-Score. As the previous one, on the x-axis, we have the score for each product according to its nutritional characteristics. On the y-axis, we have the specific content of salt. We can see that the salt content follows an almost constant curve as a function of the Nutri-Score.

On the hand, we can see that products with a high Nutri-Score tend to have a slightly lower salt content than products with a lower Nutri-Score. This analysis is logical because humans should not consume too much salt, as it is bad for health, so products didn't have a high salt content. So, there are not as many variations as for sugar.

On the other hand, there are some outliers. There is one product with a salt content of over 70g/100g, which is clearly a mistake. Humans cannot consume that much salt. In fact, this is an error. We researched the product in question, and it turns out that the salt content is 0.72g/100g.

#### Saturated fat graph

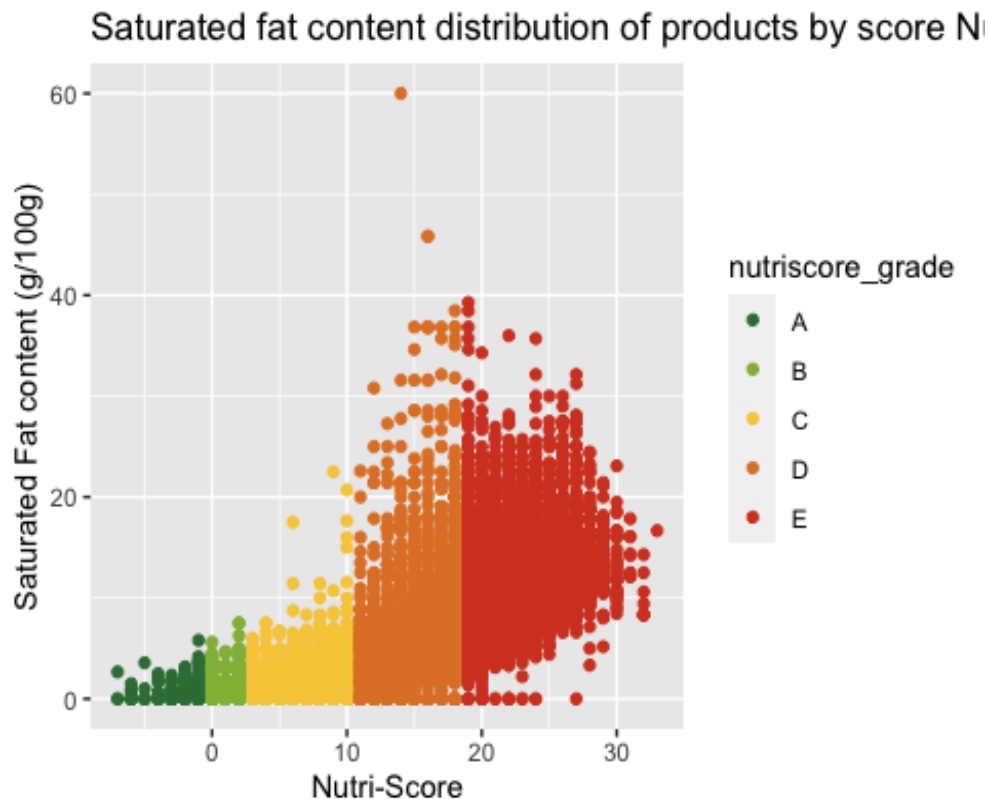
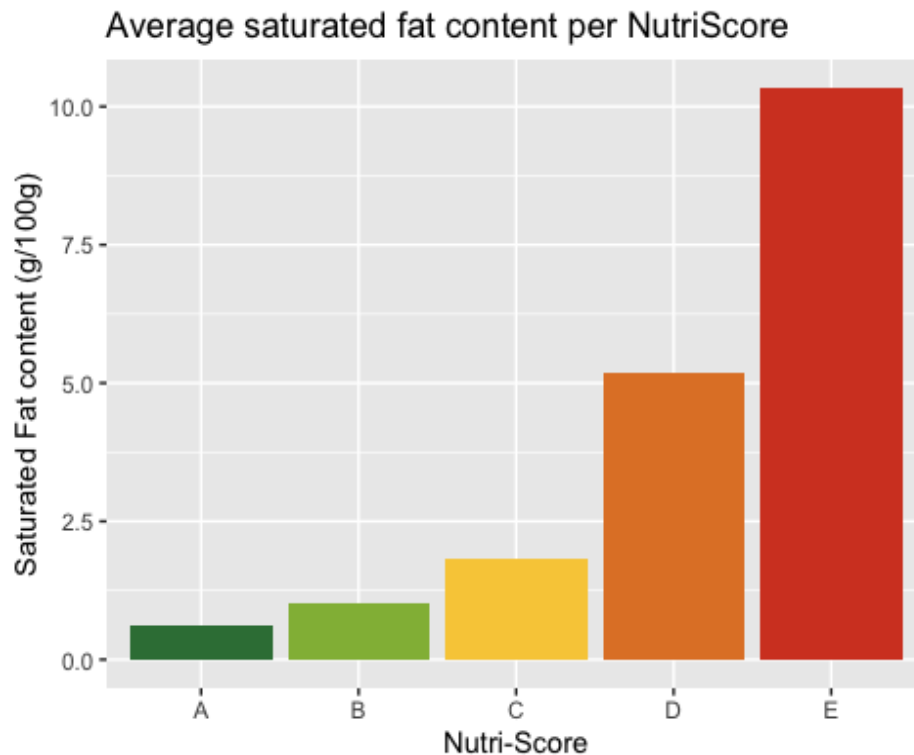


Figure 5 : Saturated fat content distribution of products by Nutri-Score



In the same way as for sugar and salt, we are going to analyze the fat content of products as a function of the Nutri-Score.

The first graph is a scatter plot showing the fat content of products as a function of the Nutri-Score. We can see that the fat content follows an almost constant curve as a function of the Nutri-Score score. However, we can see that products with a high Nutri-Score tend to have a slightly lower fat content than products with a lower Nutri-Score. This analysis is logical since the lowest Nutri-Score scores are the worst for health. Products with a score below 10 generally have a saturated fat content of less than 6g/100g, while products with a score above 10 have a much higher fat content, around 25g/100g.



The following chart shows the average fat content per Nutri-Score. On the x-axis, we have each Nutri-Score and on the y-axis, we have the mean of saturated fat content. We can see that products with a high Nutri-Score tend to have a lower average fat content. There is an increasing relationship between the Nutriscore and saturated fat content. In fact, products with a Nutri-Score A have half the fat content of a Nutri-Score B product. Similarly, Nutri-Score B products have a fat content half that of a Nutri-Score C product. Finally, Nutri-Score C products have a fat content almost three times lower than that of a Nutri-Score D product. The top of the scale is Nutri-Score E, which has a fat content almost two times higher than that of a Nutri-Score D product. This shows a clear difference between healthy and less healthy products in terms of fat content.

We can conclude from this that we should avoid eating biscuits or cakes with a nutriscore of E because saturated fat and sugar bring cardiovascular risks, increased cholesterol levels, tooth decay, overweight and obesity.

### Proteins graph

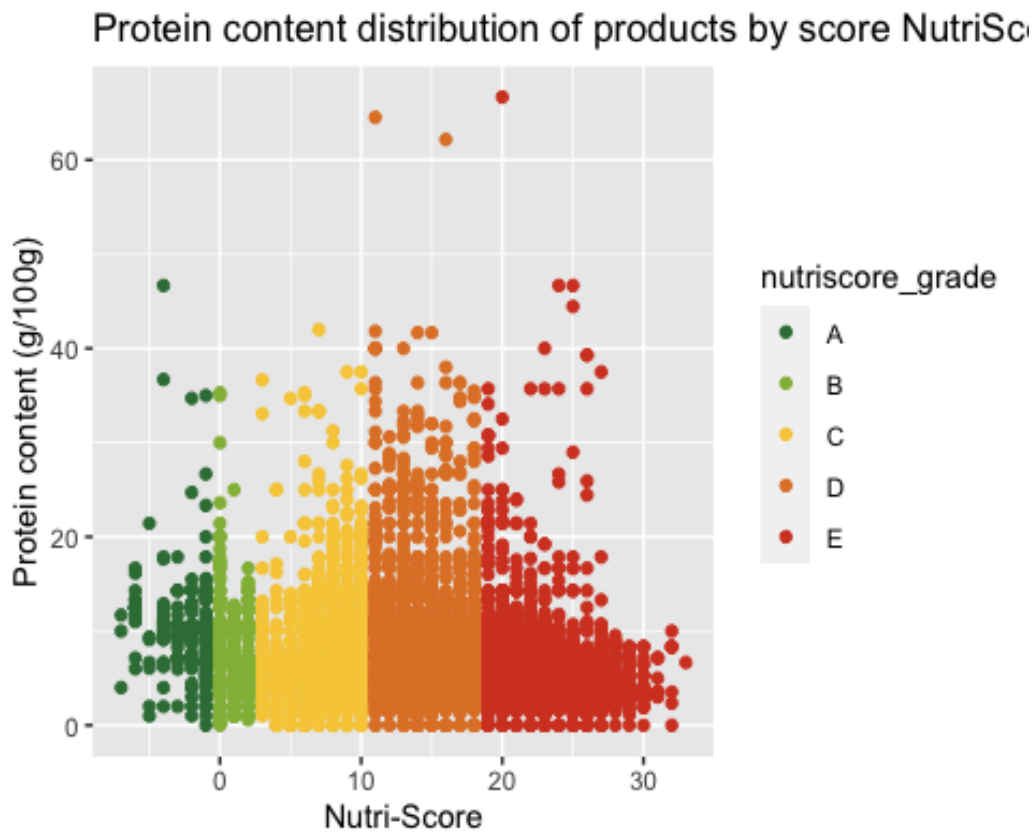
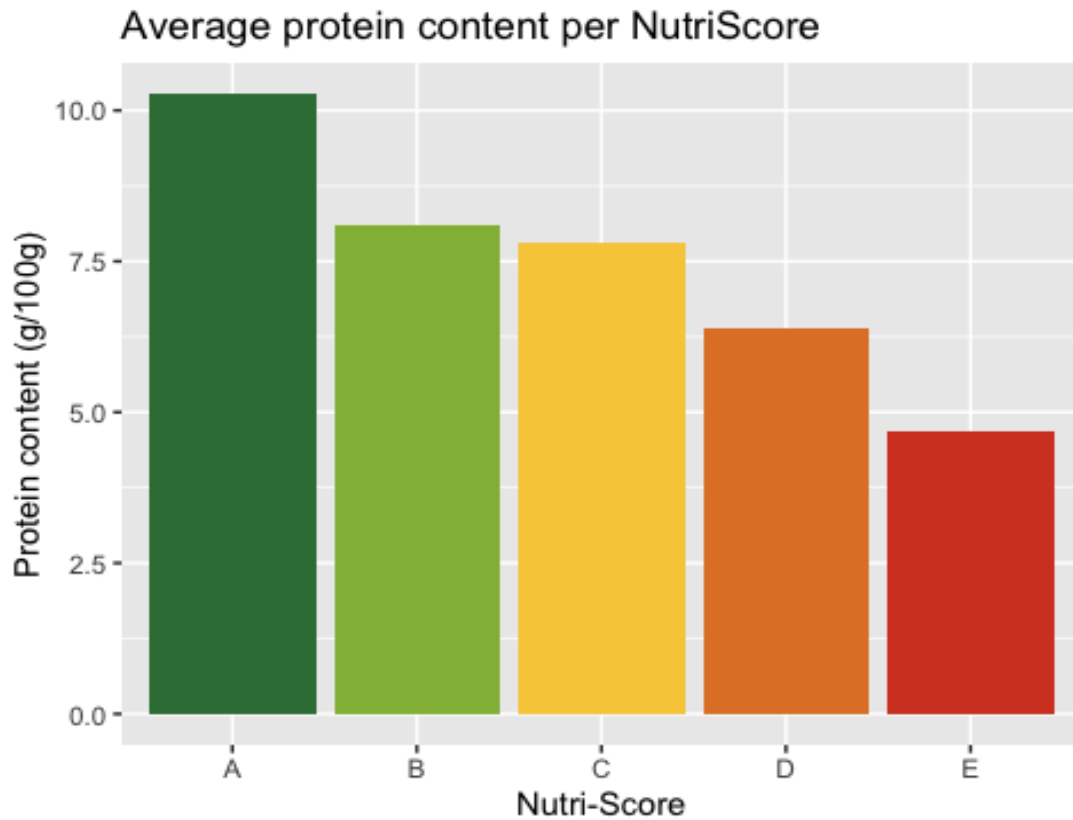


Figure 6 : Proteins content distribution of products by score Nutri-Score

The first graph is a scatter plot showing the protein content of products as a function of the Nutri-Score. Like the previous graph, the x-axis shows the score for each product according to its nutritional characteristics. The y-axis shows the specific protein content. We can see that the product data is almost equivalent for each Nutri-Score, with a high density around 10g/100g, although there is a slight drop in protein content for products with a low Nutri-Score. There are only a few points that are off, but the majority follow the constant.

Manufacturers generally opt for higher quality, more nutrient-rich ingredients, including protein sources such as pulses and wholegrain cereals. In addition, reducing less healthy ingredients, such as added sugar and saturated fat, in high Nutri-Score products can free up space for higher protein ingredients. So, the quest for a better Nutri-Score often leads to ingredient choices that favor a higher protein content in food products.

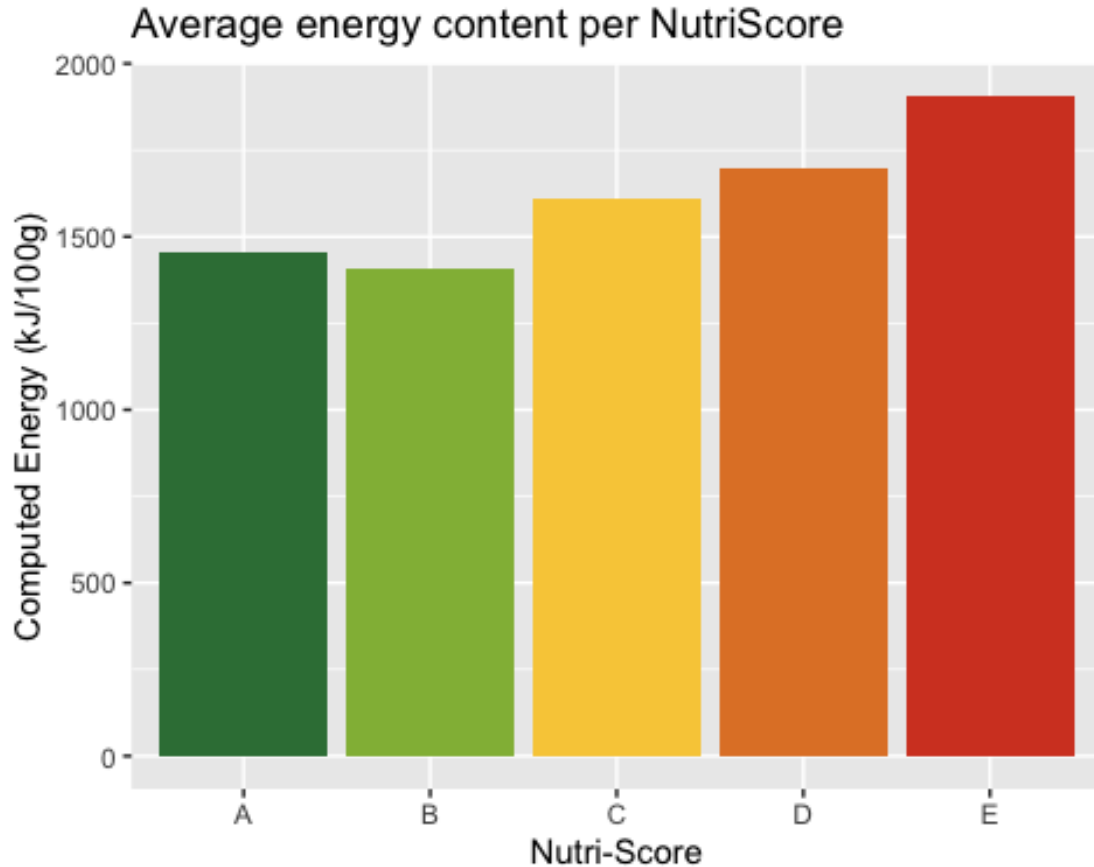


*Figure 7 : Average protein content per Nutri-Score*

The second graph is a bar chart showing the average protein content per Nutri-Score. On the x-axis, we have each Nutri-Score and on the y-axis, we have the mean of protein content.

This graph shows a clear difference between the products, as the previous graph did not show any significant difference, hence the importance of cross-referencing the data. We can see that products with a Nutri-Score A represent twice the protein content of products with a Nutri-Score E. There is a slight difference between products with a Nutri-Score B, C and D, but they are still lower than products with a Nutri-Score A. Then the product of Nutri-Score E has the lowest protein content.

## Energy graph



*Figure 8 : Average energy content per Nutri-Score*

The graph above shows the average energy content per Nutri-Score. We can see that products with a high Nutri-Score tend to have a lower energy content. In fact, products with a Nutri-Score A have an average energy content of 1500kJ/100g, while products with a Nutri-Score E have an average energy content of around 2000kJ/100g.

This shows a clear difference between healthy and less healthy products in terms of energy content. This is because products with a high Nutri-Score tend to have a lower sugar, salt, fat, and saturated fat content, which reduces the energy content of the products. There is a slight difference between products with a Nutri-Score A and B, as well as products with a Nutri-Score C and D, but products with a Nutri-Score A and B are still lower than products with a Nutri-Score C and D.

As with most of the graphs studied, products with a Nutri-Score E have the highest energy content, which is logical. The more nutriscore E products we eat, the fatter we will get, so we really need to be careful with these products, which provide too many of the calories we need to function properly.

## Boxplot Conclusion

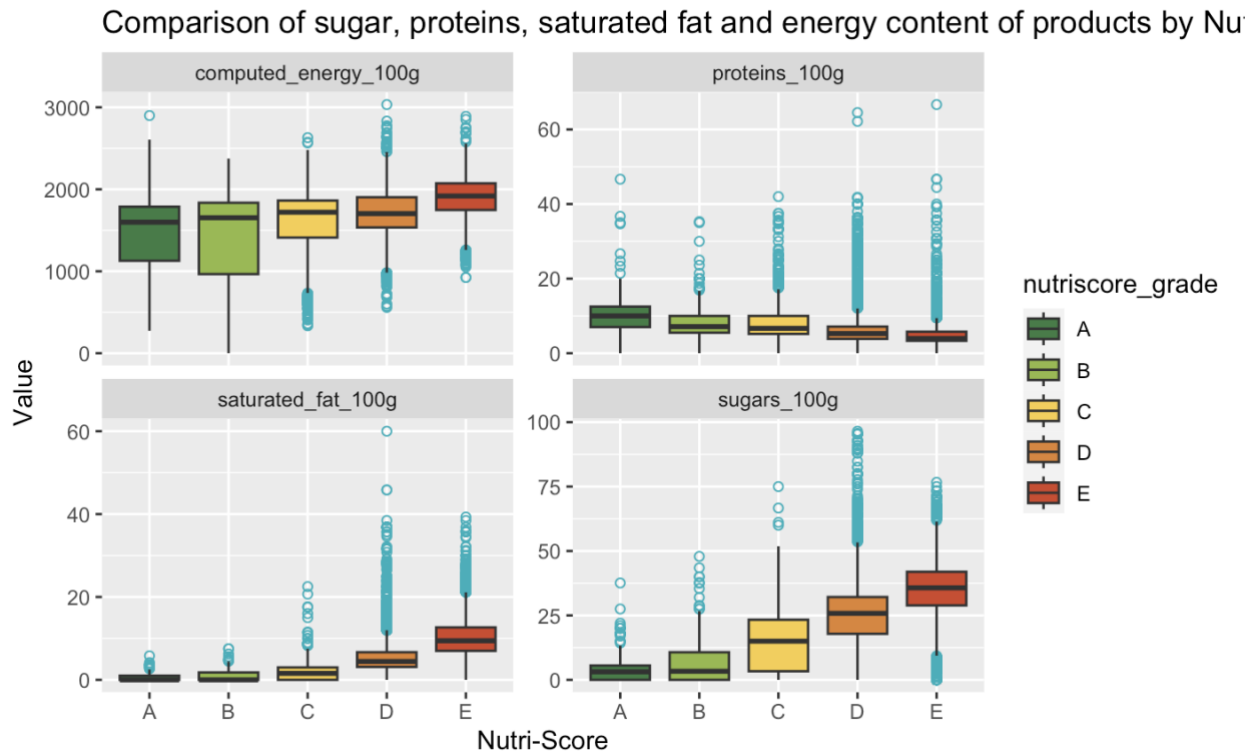


Figure 9 : Comparison of sugar, proteins, saturated fat and energy content of products by Nutri-Score

Here's a graph with several boxplots to compare the sugar, proteins, saturated fat, and energy content of products for each Nutri-Score.

This summary graph makes it easier to visualize the data. Thanks to the boxplots, we can see the median, quartiles, extreme values and outliers. We colored the outliers in blue. The median is represented by the line in the middle of the box. The box represents the interquartile range, i.e. the range between the first and third quartiles. The median is more representative because it is not influenced by the extreme values. On the other hand, these boxplots confirm the trends observed previously.

As in the previous graphs, we can see that products with a high Nutri-Score tend to have a lower sugar, saturated fat, and energy content. At the same time, they have a higher protein content. Similarly, products with a low Nutri-Score tend to be higher in sugar, salt, saturated fat, and energy. This trend is logical since the Nutri-Score is calculated based on the product's nutritional composition.

Bio, vegetarian, vegan

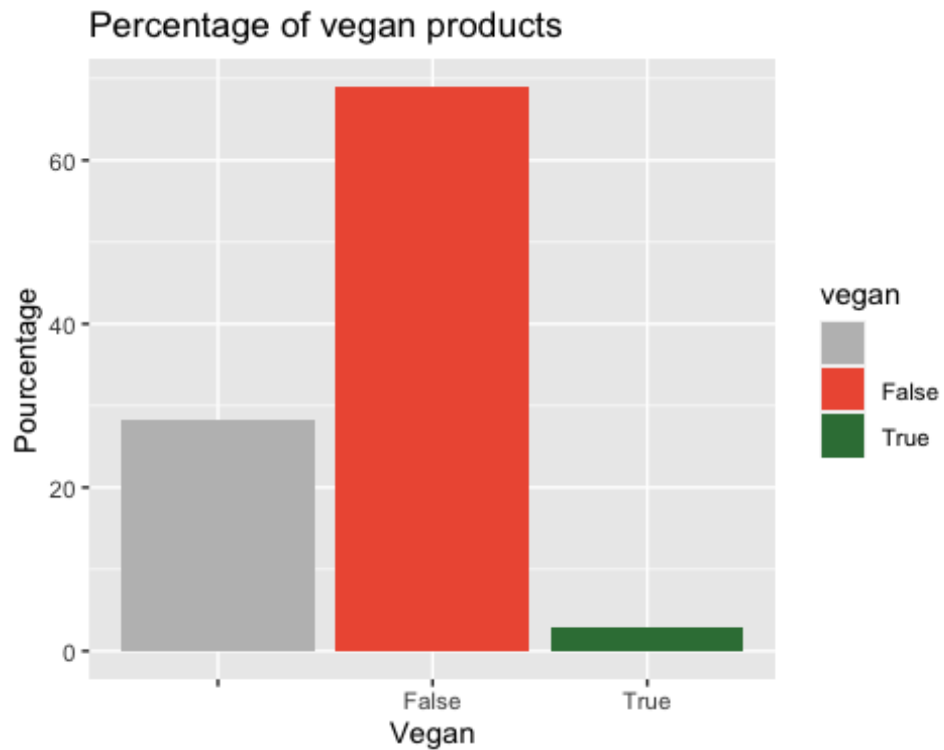


Figure 10 : Percentage of vegan products

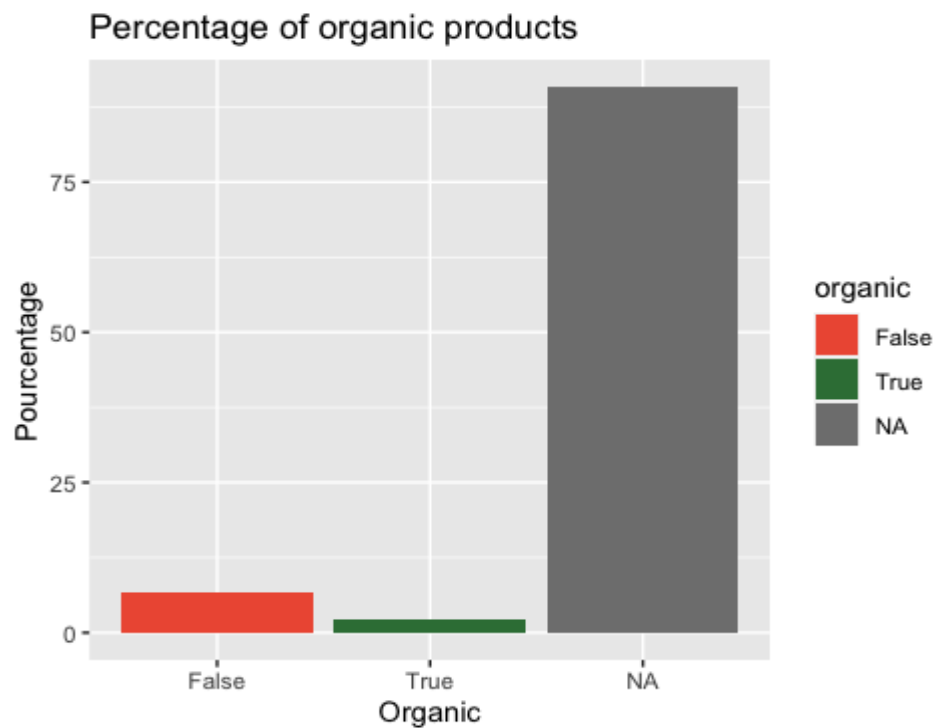


Figure 11 : Percentage of organic products

Exploiting data on the vegan biscuit category can be difficult for several reasons. Firstly, vegan products are new to the market, which means that there may be a lack of specific data available on these products. In addition, vegan biscuits often use ingredients to replace those of animal origin such as butter and eggs, which makes the nutritional data more complicated to exploit. These factors make it more difficult to analyze this category. The same conclusion applies to organic and vegetarian products.

We have therefore chosen not to use this data for our analysis. It would have been interesting to cross-reference organic, vegetarian, and vegan products with the Nutri-Score to see whether these products are healthier than conventional products. However, the data is not reliable enough to be used.

If we analyze the first graph, we can see that the part of unknown vegan products is around 28%. This is a significant proportion, which makes it difficult to analyze the data. But if we are not taking into account the unknown values, we can see that the part of vegan products is around 2%. This is a very low proportion, while the part of non-vegan products still approaches 70%.

On the second chart, the same applies to organic products, where the part of unknown products is around 82%. This is a proportion that needs to be taken even more into account, which makes it difficult to analyze the data. Because there is a clear link between the products and their nutritional characteristics and the fact that they are organic / vegan or vegetarian, but here we can't analyze it.

## Palm oil

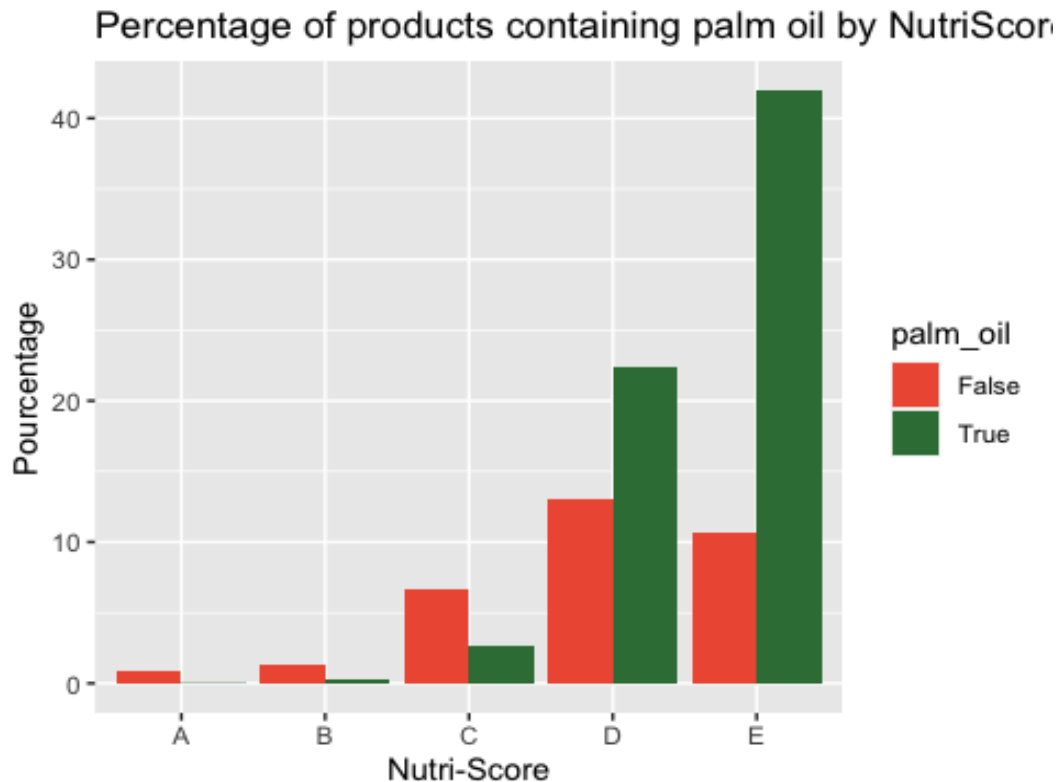


Figure 12 : Percentage of products containing palm oil by Nutri-Score

This graph shows the proportion of products made with or without palm oil, according to the Nutri-Score. On the x-axis, we have the Nutri-Score, and on the y-axis, we have the percentage of products containing or not containing palm oil. On the hand, the lower the nutriscore (close to E), the higher the proportion of products containing palm oil. On the other hand, the higher the nutriscore (close to A), the lower the proportion of products containing palm oil, or even none at all. We can also add that the proportion of products made with palm oil does not exceed that of products containing no palm oil, up to nutriscore C. From nutriscore D onwards, the proportion of products made with palm oil exceeds that of products containing no palm oil. This shows that the least healthy products are more likely to contain palm oil than the healthiest products. Finally, it is important to note that palm oil contributes to increasing the level of saturated fats in products, which is bad for health. This confirms the trends observed previously.

The price factor is also an important consideration. Palm oil is cheaper than other oils, which may explain why cheaper products are more likely to contain palm oil. This shows that the cheapest products are often the least healthy, which is a problem for public health.



### 3. Conclusion

To summary, the Nutri-Score is a good indicator of the nutritional quality of products. Our analysis is based on biscuits and cakes sold in the United States. We have seen that the majority of products in this category have a grade E and D nutriscore (87%), which means that the majority of products in this category have mediocre nutritional quality.

We have seen that products with a high Nutri-Score tend to have a lower sugar, salt, saturated fat, and energy content. On the other hand, products with a low Nutri-Score tend to have a higher sugar, salt, saturated fat, and energy content. This is because products with a high Nutri-Score tend to have a higher nutritional quality, while products with a low Nutri-Score tend to have a lower nutritional quality. This is a logical trend, as the Nutri-Score is calculated based on the nutritional composition of the product.

We also found that products with a high Nutri-Score tend to have a higher protein content. This is because manufacturers generally opt for higher quality, more nutrient-rich ingredients, particularly protein sources such as pulses and wholegrain cereals. Thus, the search for a better Nutri-Score often leads to ingredient choices that favor a higher protein content in food products.

Finally, we have seen that the least healthy products are more likely to contain palm oil. Palm oil contributes to increasing the level of saturated fats in products, which is bad for health. This confirms the trends observed previously.

In conclusion, the Nutri-Score is a good indicator of the nutritional quality of products. The Nutri-Score is calculated based on the nutritional composition of the product, i.e. the sugar, salt, fat, saturated fat, and protein content. The Nutri-Score is a color code ranging from A to E, with A being the best score and E the worst. The Nutri-Score aims to help consumers make healthier food choices.