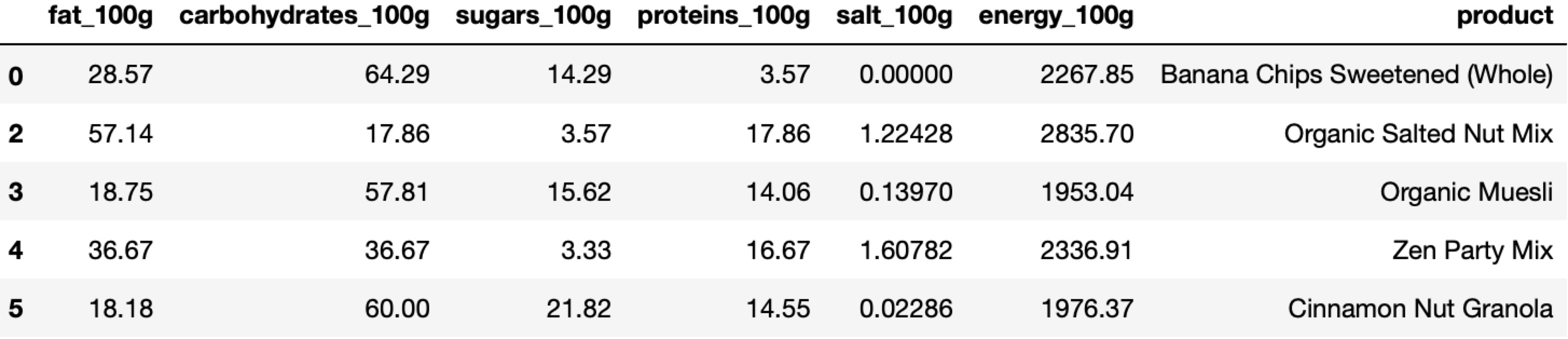
**World Food and Health-Related Issues Report**

**What is the Problem**

Nowadays, people focus more on their health. They tend to eat healthier food to improve body conditions or achieve different goals (eat fewer sugars or fat). What’s more, for specific kind of people (hypertension or diabetes), they want to choose suitable food to not only satisfy their bodies’ essential demands but also avoid making their disease worse. I, working as a data analyst, will help you frame the problem, find solutions and provide some conclusions based on analysis.

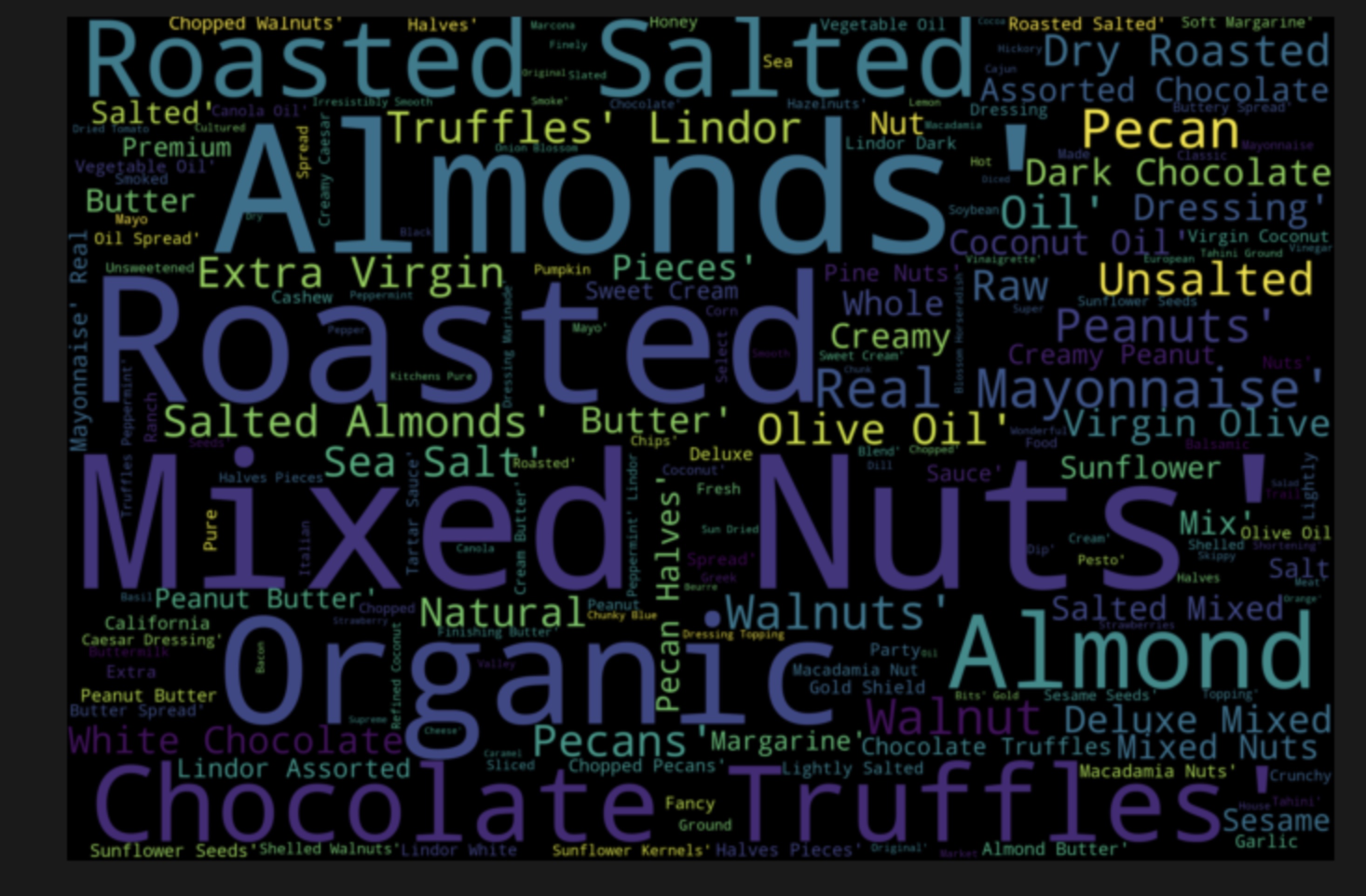
**The Process of Solving Problem**

Our data includes six main variables – fat\_100g, carbohydrates\_100g, sugars\_100g, proteins\_100g, salt\_100g, energy\_100g and one products’ column. The raw data appears as follows:



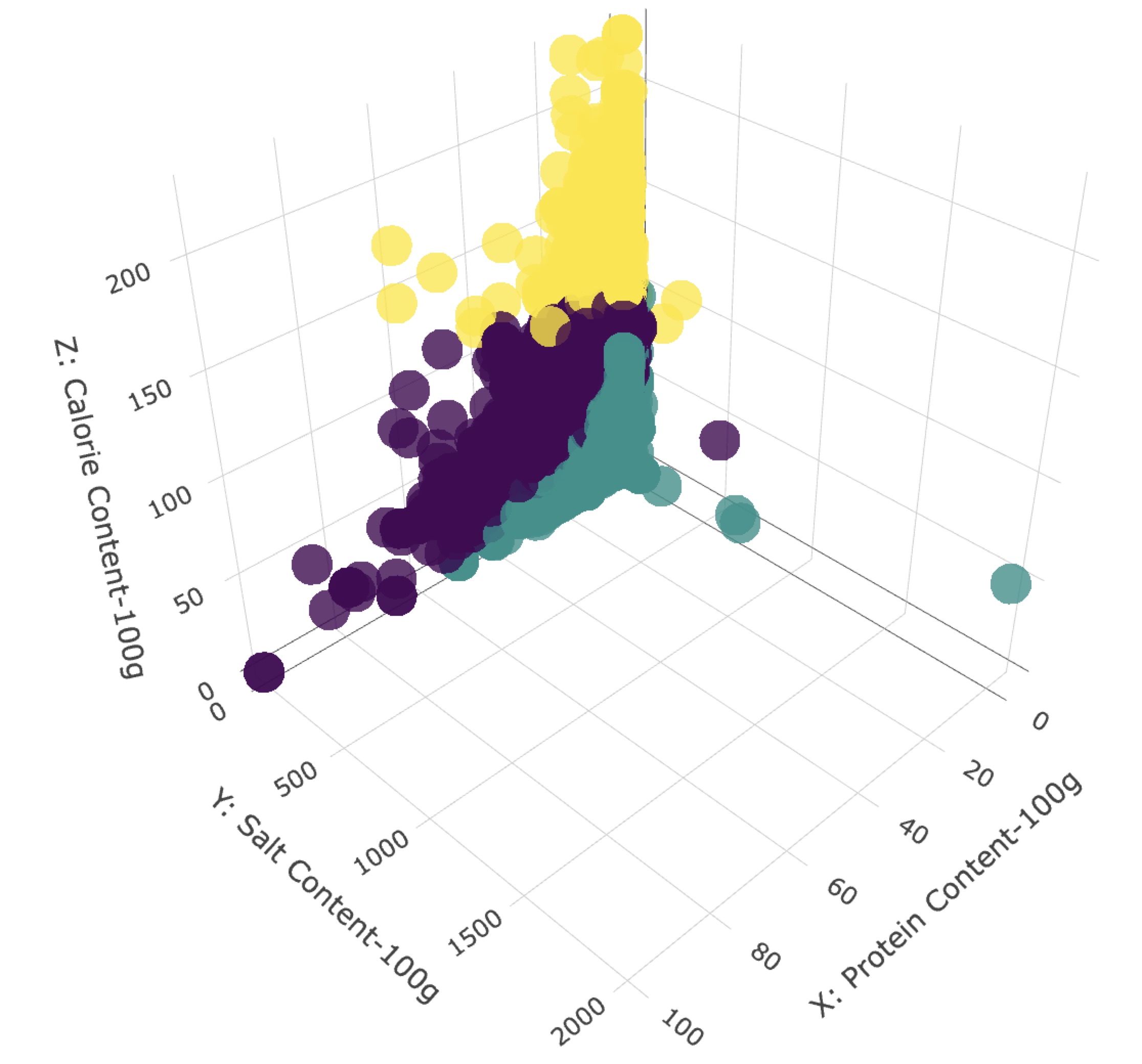
**1. Word cloud of different kinds of food**

To solve the first question – how to select the best food for meeting specific needs or avoiding eating something too much, word clouds for products based on different variables will be created. For example, for high-fat food, what kinds of products play an important part in it? More frequent of a product, bigger the size of that product name in word cloud. The high-fat word cloud is like:



**2. Clustering using different attributes**

The clustering method can help customers distinguish various food. **That original variables are not good attributes to complete products’ clustering has been proved**. Therefore, a new variable **calorie\_100g** has been created by combining fat, sugars, and carbohydrates. The plot shows it can separate food into low-calorie, average-calorie, and high-calorie.



What’s more, I conducted **PCA** to lower the dimension and produce new variables, which include **chemical element oriented, high energy, diabetes friendly, hypertension friendly**. With these new variables, customers can find suitable food for specific patients. Those kinds of food attempt to provide essential nutrition while lower the risk of making disease worse. (You can see the clustering plot based on new attributes in my notebook)

**Conclusion**

1.For people who want to lower their weight, they should eat less roasted and organic food. Also, they should consume fewer nuts, tea, candy chocolate, and ice cream. Certainly, they can find some kinds of food with lower calorie just as above clustering plot shows. For people dreaming of increasing muscles, they should ingest much more beef and various cheese. And for some office workers who want to get more energies, they can choose peanut, chocolate and organic food. These foods will help them perform better in their work.

2. For patients who get diabetes, they should eat the food with normal levels of chemical element and energies but the lower level of fat, sugars, and carbohydrates, which is combined as ‘diabetes friendly’ in the plot. The suitable range of it can be [-19, 19] (to get ‘diabetes friendly’, you should multiply respective coefficients (eigenvectors) with original data). It means that anytime you get one kind of food, if the product of its contents in 100g and coefficients is in [-19, 19], then you can eat it. Also, for patients who get hypertension, [-2, 21] is the safe range of ‘hypertension friendly’. And the way of calculating it is the same as ‘diabetes friendly’.

Kaggle: <https://www.kaggle.com/zhenyufan/unsupervised-machine-learning-of-world-food-facts>