

McDonald's Nutritional Analysis Application

Yuan Lou
ECS
University of Southampton
Southampton
yl2m17@soton.ac.uk

ABSTRACT

As the pace of life becomes faster and faster, the food culture and lifestyle have also changed, and it is followed by obesity. McDonald's is the representative of current diet culture. This article first attempts to use statistical methods to analyze the relationship between McDonald's food and health, including the relationship between food and human needs, the relationship between McDonald's food and natural food. Analysis methods include similarity, weight calculation comparison, and so on. At the same time, the implementation of WEB application helps people understand the nutritional content of McDonald's products more conveniently, and helps consumers to provide the reference for daily diet.

Keywords

McDonald's, health, similarity, diet, obesity

1.INTRODUCTION

Obesity has become a serious problem in today's society. At the same time, fast food such as McDonald's was known as a part of people's lives. This paper analyzes McDonald's products in a detailed analysis of McDonald's products, including the relationship between food and human needs, the relationship between McDonald's food and natural food. At the same time, the article describes the analysis process, starting from the requirement analysis, including data collection, design, analysis and visualization, and discussion. Finally, some conclusions were drawn about whether McDonald's is healthy. At the same time, an operational application is realized, and the user can quickly and conveniently query the analyzed information.

2.BACKGROUND

According to the World Health Organization, "the number of obese people in the world has increased nearly threefold since 1975." In 2016, more than 1.9 billion adults aged 18 and over were overweight, and more than 650 million were obese. Obesity is not only an independent disease, but also an important cause of many chronic diseases such as hypertension, cardiovascular and cerebrovascular diseases, and diabetes. [1] The underlying cause of obesity and overweight is the energy imbalance between calorie intake and calories burned. On a global scale: First, the intake of high-energy foods rich in fat continues to increase; Second, more and more forms of work such as sedentary nature, changes in transportation patterns, and increased urbanization have aggravated the lack of physical activity. [2] In other words, diet and lifestyle are the biggest causes of obesity. The closest thing to diet and lifestyle is the fast food culture, and the most well-known of them is McDonald's. McDonald's has a global presence in 119

countries on six continents. As of April 2017, McDonald's has approximately 36,000 stores worldwide and is the most well-known brand in the global catering industry (drinking). For consumers, McDonald's brings convenience and speed to people, but it also seems to have caused incentives for people's obesity, because there are reports that some products of McDonald's, such as chicken burgers, may contain ingredients and fats that consumers never know[3]. McDonald's has hinted that its products are healthy and can be eaten every day. However, consumers still cannot directly judge whether McDonald's products meet the needs of the human body.

3.REQUIREMENT ANALYSIS

Based on the above background, as a consumer, it is inevitable to ask: Does McDonald's products lead to obesity? What are unhealthy foods in McDonald's? How much do we eat to meet our human needs? Can McDonald's products be equated with natural foods?

My challenge is to solve the above problems. So I'm trying to build a data application to help consumers better understand the nutritional components of McDonald's products and to remind them of their daily dietary intake records.

3.1.Requirement of Analysis

Table 1. Requirement of Analysis

	Analysis
A1	Analysis of the nutritional components of McDonald's, correlation, the highest nutrition, the lowest nutrition
A2	To analyze and compare the relationship between McDonald's food and human needs, and classify McDonald's food into high calorie, high fat and unhealthy foods.
A3	Analysis and comparison of the relationship between McDonald's food and common natural foods, fruits and vegetables

3.2.Requirement of functions

Table 2. Requirement of functions

	functions
F1	The application screens beverages, hamburgers and meat according to product type.
F2	Screening for high calorie and fat food based on nutritional types
F3	Visualization of Nutrition Distribution List of McDonald's Products
F4	Visualization of the relationship between McDonald's products and natural foods

3.3.Requirement of no-functions

Table 3. Requirement of no-functions

	No-function
N1	Friendly UI Design
N2	Responsive layout to facilitate mobile tablet browsing

4.DESIGN

As a data science application, it must include appropriate data analysis and completed software engineering practices. The application must cover the process of data collection, data processing, analysis and visualisation. Application development must follow software engineering specifications to meet industry standards. According to the requirements analysis, the data analysis tool uses python because it provides a rich package and the syntax is simple.

For the process of analysis methods, the basic methods in statistics are necessary, including similarity, weight calculation comparison, etc.

For software function design, it must meet the ease of use and integrity, try to use some languages and tools, JS, ajax, bootstrap, considering the data is not complicated and the reading speed, you can choose the json format. Visualization of the application is D3, echart.

5.DATA COLLECTION AND CLEANUP

5.1.Data collection

Collecting the right data is the first step. According to the needs analysis A1A2, McDonald's nutrition data set is a must, in order to compare and human requirement, I need the human DRI data set, and finally I need nutritional data containing some natural foods.

The data collection process is not complicated, and McDonald's data can be obtained from kaggle. They are from kaggle and openfoods, both of which are open source data. Regarding the food data set, I queried the openfoods website. After screening, I collected five data that can be used, namely McDonald's nutritional data, human minimum requirements, human requirements, foods ingredient data, and another food data sets.

5.2.Data observation

The McDonald's data set contains information on the nutritional content of 260 McDonald's products, including size, calories, fat, cholesterol, etc., as well as the proportion of daily needs, the data is basically complete.

The food data set contains nutrients for more than 92,000 foods, including vegetables, fruits and grains. Some of the missing values appear.

Additional food data sets contain information on the nutritional content of 59 common foods, including vegetable and fruit snacks, with some missing values.

The daily nutrient requirements of the human body, including minimum and standard values.

5.3.Data cleanup

For the McDonald's data set, the McDonald's data set is complete after missing values. In order to achieve analytical demand A3, I hope to screen out suitable matching foods, focusing on healthy natural foods such as fruit and vegetable staple foods. Therefore, the second processed food such as fried chicken pizza coffee is primary removed. For the selection of eigenvalues, the focus is on the six major nutrients the body needs, sugars, oils, proteins, water, inorganic salts and vitamins. [4] I also unify the eigenvalue names of all datasets, and the unit is unified. For example, calories are chosen Kcal instead of J.

After data cleansing and screening, three preliminary processed data sets were obtained. One is the McDonald's Nutrition Data Set, which contains 260 McDonald's food ingredients and the percentage of human needs. The other is the natural food nutrition data set. There is also a data set of human intake and minimum human intake.

6.BASIC ANALYSIS OF MCDONALD'S

6.1.The relevance of McDonald's nutrition

In order to study the relationship between nutrition, I tried to calculate the correlation of nutrition. For more in-depth research, I tried to calculate the Pearson correlation coefficient, which is a measure of vector similarity. The output range is -1 to +1, 0 means no correlation, negative is negative correlation, positive value is positive correlation. Pearson's formula is:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

After visualization I see the relevant map, figure1 . As can be seen from the correlation diagram, the more clearly the yellow represents the higher the positive correlation, we can already see the features that are clearly interrelated (more in the yellow part of the figure), and the negative correlation is black and dark blue. After some observations, I found that the relationship between calories, fat, and cholesterol is positively related, that is, they often appear in food at the same time.

However, there are some surprising correlations that are not well understood. For example, the correlation between total fat and saturated/trans fat is rather weak, and as consumers, it may be considered that there is a strong correlation between them. The heat map also presents interesting findings from the spots of the negative correlation map (dark blue/black). For example, it indicates that carbohydrates are usually negatively correlated with trans fats, cholesterol, sodium, dietary fiber, and vitamin A.

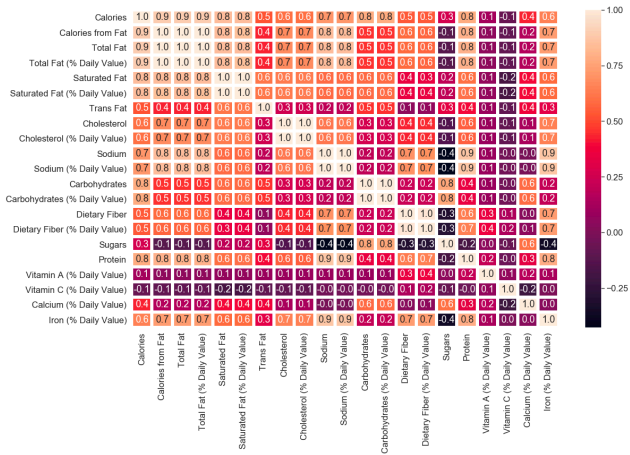


Figure 1. The relevance of McDonald's nutrition

6.2.Nutrient distribution

I used basic statistical methods to calculate the nutritional distribution of McDonald's, including the highest and lowest nutritional items and menus.

Table 4. Highest Item

Highest	Item
Calories	Chicken McNuggets (40 piece)
Calories from Fat	Chicken McNuggets (40 piece)
Total Fat	Chicken McNuggets (40 piece)
Cholesterol	Big Breakfast with Hotcakes (Regular Biscuit)
Sugars	McFlurry with M&M's Candies (Medium)

Table 5. Lowest Item

Lowest	Item
Calories	Diet Coke (Small)
Calories from Fat	Side Salad
Total Fat	Side Salad
Cholesterol	Hash Brown
Sugars	Hash Brown

Table 6. Highest Menu

Highest	Category
Calories	Chicken & Fish
Calories from Fat	Breakfast
Total Fat	Breakfast
Cholesterol	Breakfast
Sugars	Smoothies & Shakes

Table 7. Lowest Item

Lowest	Category
Calories	Beverages
Calories from Fat	Beverages
Total Fat	Beverages
Cholesterol	Beverages
Sugars	Snacks & Sides

7.MCDONALD'S NUTRITION AND HUMAN NEEDS

After the above analysis, we probably learned about the nutritional status of McDonald's products, and also learned about high-nutrition and low-nutrition products, but this product is relative to McDonald's, for the daily needs of the human body, they are still High nutritional food? How does McDonald's meet the nutrients needed for a day? So I tried to introduce more detailed analysis of the daily demand data of the human body, and I hope to pick out some unhealthy foods relative to the human body.

7.1.McDonald's classification

In order to remind users and people who are losing weight, I hope to mark food as high-calorie food, high cholesterol, high fat, and unhealthy food.

For the daily diet, three meals a day must meet the nutritional needs of the human body for a day, so on average, the nutrition of each meal should meet one-third of the body's needs. Therefore, my data standard is 30%, which means that more than 30% of the daily demand for food is in line with high nutrition.

At this time I introduced the human needs data set. After entering the comparison with the McDonald's data set, I have set a point for high calorie >900Kcal, high cholesterol >90mg, high fat >30g. If a food is labeled as high calorie, high cholesterol and high fat, it is marked as unhealthy.

The method I analyzed is to use the functions in the Python language, and finally count the number of products that meet the standard. The following figure visualizes the statistics of some products that meet the above conditions. More than 30% are marked as red, and at the same time in the data set. Mark, update the data set. After statistics, there are 6 high-calorie foods, 29 high-cholesterol foods, 29 high-fat foods, and 4 unhealthy foods. After observation, it is surprising that most of the high-cholesterol and high-fat foods are breakfast, and there are many beef and pork categories. Usually high cholesterol and high fat occur at the same time, which is also consistent with the above analysis of nutritional relevance. Unhealthy foods are the regular and large sizes of Big Breakfast with Hotcakes, and Chicken McNuggets 20, 40. It can be seen that these two kinds of foods will provide excess calorie cholesterol and fat to the human body in one meal, and the weight loss group should avoid it.

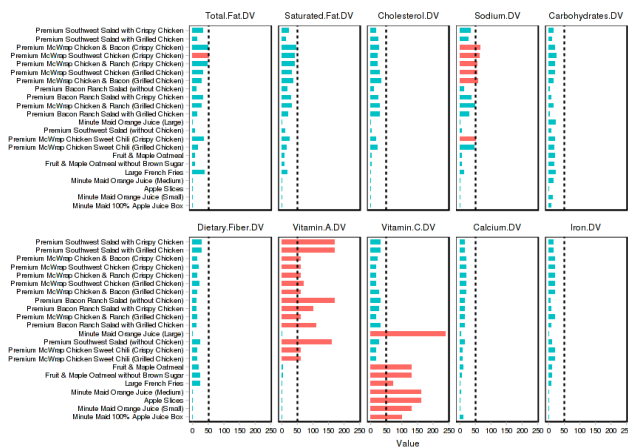


Figure 2. The relevance of McDonald's nutrition

7.2.How many McDonald's meet the needs of the day?

I am trying to make an interesting analysis. If I eat McDonald's for three meals a day, how can I meet the needs of the human body?

First of all, I first quoted the daily minimum demand data of the human body. Taking an adult as an example, after inquiring, the minimum calorie intake per day for adults is about 2000, so the average food intake per meal should be 600-700 calories.

Breakfast: I sorted out the calorie-compliant food from the breakfast menu, Bacon, Egg & Cheese Bagel for 620 kcal

Lunch: The menus for lunch are Beef & Pork, Desserts, Beverages, Coffee & Tea, Smoothies & Shakes, because lunch can't be a single product, and drinks and snacks are needed, so the three calories are selected to meet the standard food. The final products are Beef Hamburger, Oatmeal Raisin Cookie and Mango Pineapple Smoothie (Small), with a total calorie capacity of 600 calories.

Dinner: After the above menu nutrition analysis, the dinner selection is lighter and the product is rich. Chicken & Fish, Salads, Snacks & Sides, Beverages. After screening, Chicken McNuggets (6 piece), Premium Bacon Ranch Salad (without Chicken), FatFreeChocolateMilkJug, Kids French Fries, their total calories are 660 calories to meet demand.

8.COMPARISON OF MCDONALD'S AND NATURAL FOOD

After the above analysis, users have already understood the nutritional value of McDonald's products, but at this time consumers can not directly remember the nutrition of McDonald's products when purchasing, because as a secondary processing product, it is impossible to evaluate the nutritional distribution through appearance and experience. So I tried to compare McDonald's products with natural foods to help people remember the nutritional value of McDonald's products through the nutritional value of natural foods.

8.1.Data selection

According to demand A3, in some product categories, I screened out three non-processed products, respectively, Vegetables and Vegetable Products, Fruits and Fruit Juices, Nut and Seed Products.

For the selection of eigenvalues, the six major nutrients, sugars, oils, proteins, water, inorganic salts and vitamins required by the human body are preferred. There is also cellulose. It has been observed that the McDonald's data set does not contain water information and total vitamin content, and the food data set does not contain information on inorganic salts.

Because each product is relatively independent for missing values, there is no linear relationship between variables, and it is not suitable for filling with regression methods, but we can find that certain rules can be found between elements with higher correlation. Such as calories and protein, calories and nano, calories and Saturated Fat. So I can calculate the vacancy value according to the ratio between the elements, and refer to the distribution of elements similar food, such as Apricots with skin and Apricots without skin. finally eliminate outliers.

After sorting out the data, I see the final food data set with six characteristic values: calories, protein, dietary fiber, Sugars, Sodium, and Saturated Fat.

8.2.Nutrition score and match

In order to judge the nutritional relationship of food, I hope to carry out the same amount of comparison or multiple calculation, so the value is the best standard, so that each product has a nutritional score calculation. Formula:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6$$

a_i represents the weight of each feature; x_i represents the feature; y is the corresponding commodity, that is, the score. Because the product contains six characteristic values, calories, opal, dietary fiber, and sugar are essential nutrients for the human body, so the weight is 20% each. Sodium, Saturated Fat is lower for the daily needs of the human body, so the weight is 10%. So the final formula is:

$$Y = 0.2X_1 + 0.2X_2 + 0.2X_3 + 0.2X_4 + 0.1X_5 + 0.1X_6$$

After calculation, I got the nutritional score of McDonald's and natural food. This score reflects the nutritional level of the food and is comparable. The equal and multiple relationship between the scores can also represent a match between the food and the food. According to the calculation, the match with too many multiples is discarded, and finally the final data set is obtained. The table shows some of the matching results, and the full matching results can be queried in the app.

Table 8. Score and Match

Item	Score	Match
Big Breakfast (Large Biscuit)	337	2*Eggplant, pickled
Strawberry Shake (Large)	226	Celery flakes, dried
....
Hotcakes and Sausage	204	2*Nuts, almond paste
Baked Apple Pie	72	Currants, zante, dried

9.APPLICATION IMPLEMENTATION

After designing and analyzing, I got the final result, which included the classification of McDonald's food and the match with the food. Based on the above functional requirements analysis, my goal is to allow users to quickly query previously published work, i.e., what is the unique contribution that your work makes to the field? Please consider what the reader will learn from your submission, and how they will find your work useful. If you write with these questions in mind, your work is more likely to be successful, both in being accepted. So I making an application (Web) to show my result, which is a relatively difficult process. My website uses the front-end technology mainly with js replaced the back-end tech such as php, so the website more flexible. The technology used js, ajax, css, while using some frameworks and libraries such as bootstrap, echarts.js and so on. It is worth mentioning that, almost all of the site using js, ec6 syntax. So

node.js is an essential environment for programming and requires tools like gulp and babel to support this syntax. So to say, I create a js plugin with js, all of page is provided by this plugin. For data storage, the traditional database has not been used, but instead of json file, it is more effective to interact with js.

Table 9. Functions Test

	functions	Test
F1	The application screens beverages, hamburgers and meat according to product type.	Pass
F2	Screening for high calorie and fat food based on nutritional types	Pass
F3	Visualization of Nutrition Distribution List of McDonald's Products	Pass
F4	Visualization of the relationship between McDonald's products and natural foods	Pass

Table 10. No-Functions Test

	No-function	Test
N1	Friendly UI Design	Pass
N2	Responsive layout to facilitate mobile tablet browsing	Pass

10.LIMITATION

This application completes the analysis and visualization of the data first, but there are some shortcomings. When the data is missing missing values, the results of other data sets are not referenced, or the gap values are calculated based on the correlation between the elements. For the selection of eigenvalues, the six major nutrients of the human body are important, but other nutrients also occupy the position. It is better to select the eigenvalues based on the scientific data of the body's absorption of nutrients. When classifying McDonald's products, my standard is based on three meals a day, that is, 30%, but the body's demand for each nutrient is different, and the human body will produce excrement and nutrient consumption, which should also be used as The criteria to consider. Similarly, when calculating the nutrient score, the choice of weight should use a more scientific basis. If there is a data set, consider training weights through machine learning.

11.CONCLUSION

This paper realizes a scientific data application through the analysis of McDonald's data. This article begins with the analysis of requirements to describe the analysis process, including data collection, data observation, and data cleansing. The analytical methods were then developed to include methods for nutritional relevance, McDonald's and human relationships, and relationships with natural foods. Finally, some conclusions were drawn about whether McDonald's is healthy, and the matching and nutritional scores between foods were achieved. Through this application, users can easily query the required information and quickly plan their own diet plans.

12.REFERENCES

1. Fund W C , Eng D C . Food, nutrition and the prevention of cancer: a global perspective[M]// Food, nutrition and the prevention of cancer :.Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.
2. Steppan C M , Bailey S T , Bhat S , et al. The hormone resistin links obesity to diabetes.[J]. Nature, 2001, 409(6818):307-12.Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.
3. Steppan C M , Bailey S T , Bhat S , et al. The hormone resistin links obesity to diabetes.[J]. Nature, 2001, 409(6818):307-12.Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.
4. Nutrition C O A . Nutrient Requirements of Dairy Cattle[M]// Nutrient requirements of dairy cattle. National Academy of Sciences, 2001.