

Accuracy of McDonald's Caloric data*

Are the calorie and nutritional value claims of fast food restaurants as accurate as they claim?

Aamishi Avarsekar

April 19, 2024

The vast availability of fast-food could lead to compromised health due to cheap prices. This paper analyses food nutrition and estimates the calories in McDonald's menu items using a linear model. It is found that these values are close to what the company claims. This accuracy can help customers make informed decisions about their health.

Table of contents

1	Introduction	2
2	Data	2
2.1	U.S. Department of Agriculture (USDA) (Legacy Food Data) (USDA FoodData Central (2024)):	2
2.2	MenuStat.org's McDonalds Data (MenuStat.org (2024)):	3
2.3	Data Measurement and Limitations:	5
3	Model	5
3.1	Model set-up	6
3.1.1	Model justification	6
4	Results	7
5	Discussion	7
5.1	Weaknesses and next steps	7
	References	8

*Code and data are available at: <https://github.com/aamishi/McDonaldsCaloriesPrediction>

1 Introduction

With the boom of fast-food restaurants in the USA, it has become easier for people to purchase more food at a lesser cost. Food insecure households and people with low incomes rely on cheap fast-food options to feed their families as compared to grocery shopping for fresh foods and preparing meals at home. More often than not, the main consumers of these fast-food restaurants are more concerned with having filling meal over a nutritious meal. While the menu items in places such as food-food restaurants such McDonald's may provide cheap food options for the quantity of the food they provide, we should also concern ourselves with their claimed nutrition values.

Customers rely on these facts that are provided by the companies themselves and it is virtually impossible to find out how many calories a small packet of potato fries actually has. People can watch what they eat for several reasons including general health reasons, personal body building goals or tracking their energy intake from different food groups.

The motivation behind this paper is to see if the fast-food company McDonald's' menu items' calories are accurate to help people take informed decision about their nutrition. The selling point of fast-food restaurants is their cheap prices, so people often tend to overlook their nutritional packing and average caloric value. With the chronic increase of fast-food availability, people may not realize how easy it is to overeat these foods. On the other hand, calorie restricting persons such as body builders prepping for a competition or people suffering from eating disorders, can also gain a sense of confidence in the nutritional values of the foods the company provides.

I built a linear model using the amount of protein, carbohydrates, fat, fiber, sugar, sodium and cholesterol as predictors to estimate the number of calories in the menu items. I trained this model on the general foods nutrition data and then applied it to the McDonalds' menu items to verify the accuracy of their claimed calories. The R Programming Language (R Core Team (2023)) was used to obtain and clean the necessary datasets and create the model.

2 Data

2.1 U.S. Department of Agriculture (USDA) (Legacy Food Data) (USDA FoodData Central (2024)):

The U.S. Department of Agriculture (USDA) (USDA FoodData Central (2024)) maintains a food database of all foods that are available for Americans to consume. This database includes a comprehensive list of nutritional values of foods of different origins. The main sources categories include foundational foods which comprise of whole foods with no processing such as fruits, vegetables and meat. The National Health and Nutrition Examination Survey (NHANES) of America records an extensive survey of people and their average diets, caloric

consumption, macro-nutrient consumption and other health vitals like cardiovascular health. The food obtained from this survey is recorded in USDA’s Food and Nutrient Database for Dietary Studies. However, the dataset of my interest was the Legacy food data collection. This dataset comprises the values of my predictors of interest from foods that have been collected through decades worth of research, published literature, and both private and public food composition databases. Although retired in 2018, this dataset was most apt for my paper because it contains cumulative data of private and public food sources which also includes fast food companies. This would help train the linear model better, as it also represents the type of fast foods that we are interested in. The trade-off here in the lack of information after 2018, as this database was transitioned to be maintained with the foundational and branded foods database respectively. However, for the purposes of training this model, the data size was at a much more handleable size, with the biggest file being around 36.4MB as compared to the biggest file of the branded food dataset with a size of 1.73GB.

For training the model, I selected the variables `calories`, `gram_weight`, `protein`, `carbs`, `fats`, `fiber`, `sugar`, `sodium`, `cholesterol`. The serving size in grams of each food item was extracted and then inner joined with the nutritional data against the `food_id` variable. The following table is quick look at the USDA Legacy food data after it was cleaned and merged to produce appropriate variables.

Table 1: Nutritional Value of Foods According to USDA

food_id	gram_weight	calories	protein	carbs	fat	fiber	sugar	sodium	cholesterol
167512	34	307	5.88	41.18	13.24	1.2	5.88	1059	0
167513	44	330	4.34	53.42	11.27	1.4	21.34	780	0
167516	38	273	6.58	41.05	9.22	2.2	4.30	621	15
167516	39	273	6.58	41.05	9.22	2.2	4.30	621	15
167517	33	309	7.42	48.39	9.49	2.6	4.41	710	13
167517	28	309	7.42	48.39	9.49	2.6	4.41	710	13

Nutritional Value of Foods According to USDA

2.2 MenuStat.org’s McDonalds Data (MenuStat.org (2024)):

Similarly, the nutritional information of menu items at a typical McDonalds’ in the USA was obtained from MenuStat. MenuStat maintains a free nutrition database of foods that are found in America’s top 100 largest chain restaurants. This database is updated annually and made available to the public via download. The website manually collects nutritional information that are made available through different media like rest websites and applications. Any missing information is supplemented for using other formats of nutritional facts such as nutritional tables. A possible shortcoming of this website is that it does not have the nutritional

information of every food item that could be available on the website. This website does not claim for the accuracy of these values. It simply provides a single public source of nutritional facts that are provided by the respective rests.

For this paper, I will be using the 2022 version of this data. The data is provided as an excel sheet in the `.xlsx` format. From this data, I was interested in similar variables as the general foods nutritional data to use to test the model. I extracted the following variables after filtering for McDonald's in the 'restaurants' column.

Table 2: Nutritional Value of McDonald's According to MenuStat

food_id	food_name	gram_weight	calories	protein	carbs	fat	fiber	sugar	sodium	cholesterol
174	Hamburger	100	250	12	31	9	2	6	520	25
175	Cheeseburger	114	300	15	33	12	2	6	750	40
176	Double Cheeseburger	165	440	25	34	23	2	7	1150	80
178	Quarter Pounder with Cheese	198	510	29	40	26	3	9	1190	90
180	Big Mac	214	540	25	45	29	3	9	1040	75
182	Double Quarter Pounder with Cheese	279	740	48	40	42	3	9	1380	155

Nutritional Value of McDonald's According to MenuStat

To understand what the values look like, we can take a look at the calories for some of the menu items:

Using `as id variables`

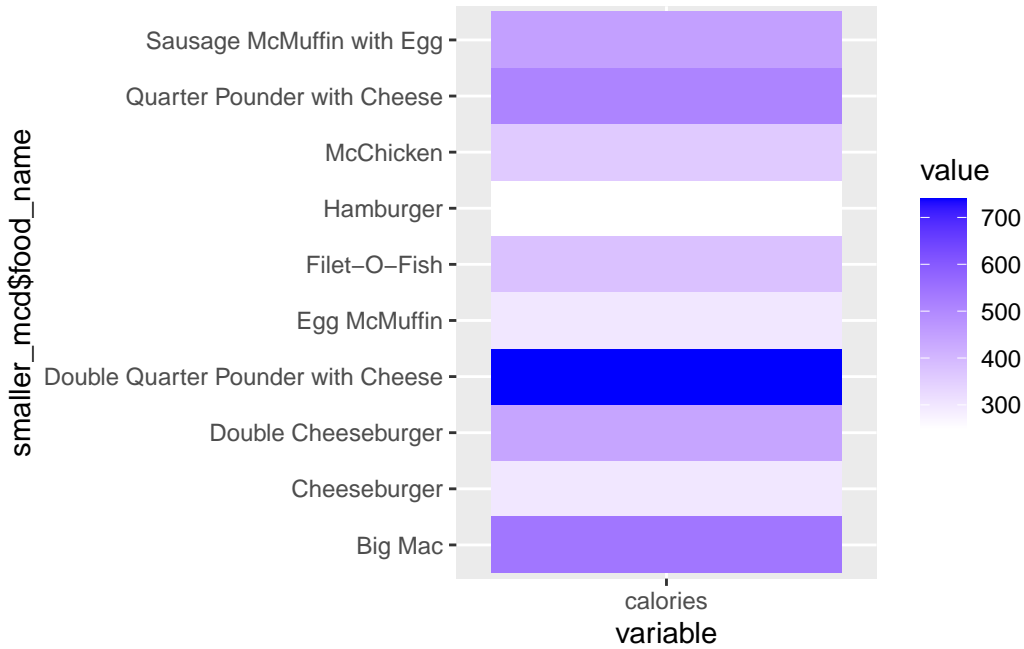


Figure 1: Nutritional Value of McDonald/“”s According to MenuStat

2.3 Data Measurement and Limitations:

The primary variable of interest in the number of calories in food. This value is measured in KCals which is the amount of energy that is required to raise the temperature of water by one degree Celsius. The rest of the variables are measured in grams. These values are then compiled over several foods to produce the database.

The USDA data on food nutrition is a compilation of scientific research and publications. The McDonald's data from MenuStat.org is also simply a compilation of the nutritional facts found on the company's websites and applications. The accuracy of this data is not guaranteed by the websites that maintain it.

3 Model

The goal of this model is to use the USDA Legacy food data as training data to establish a linear relationship with calories and other independent variables like protein, carbohydrates, fats, sugar, sodium and cholesterol. This model is generated using the R programming language's (R Core Team (2023)) `lm()` function.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

The model used to estimate the calories in McDonalds' menu items is represented by the following equation:

$$\begin{aligned}\text{calories} = & \beta_0 + \beta_1 \cdot \text{gram weight} \\ & + \beta_2 \cdot \text{protein} \\ & + \beta_3 \cdot \text{carbohydrates} \\ & + \beta_3 \cdot \text{fats} \\ & + \beta_4 \cdot \text{fiber} \\ & + \beta_5 \cdot \text{sugar} \\ & + \beta_6 \cdot \text{sodium} \\ & + \beta_7 \cdot \text{cholesterol} + \epsilon\end{aligned}$$

Here,

- calories: is the dependent variable that we want to predict for McDonald's menu based on USDA's Legacy Food data.
- `gram_weight`, `protein`, `carbohydrates`, `fats`, `fiber`, `sugar`, `sodium`, `cholesterol` are the independent variables of the model

3.1.1 Model justification

The model choice for estimation is the number of calories found in McDonald's fast food because it is an area of growing concern regarding a person's general health. Despite the size or quantity of the food that a person may be able to purchase, there is no way they can estimate the number of calories the food has. Training our data on decades old data ensures that all food types, their quantity and nutritional values has been captured. A known discussion in the body building community of finding foods that are low in calories, high in protein and low in fats and carbohydrates, goes to show that the quantity of food has no direct relationship with the number of calories that it has. We must instead rely on the calories that are derived from different nutrient types.

4 Results

5 Discussion

5.1 Weaknesses and next steps

Weaknesses and next steps should also be included.

References

- MenuStat.org. 2024. *MenuStat.org*.<https://www.menustat.org/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- USDA FoodData Central. 2024. *USDA FoodData Central*. <https://fdc.nal.usda.gov/>.