**Is fast food really unhealthy?**

**Abstract:**

In this paper, we will apply machine-learning to test and determine whether fast foods are unhealthy using popular fast foods nutrition data, and geo-coded fast food prevalence data from multiple sources. To define healthy and unhealthy, we will first perform statistical analyses on nutrition data to compare with recommending nutrition intake issued by FDA and present the results with descriptive statistics. However, instead of just labeling everything with either healthy or unhealthy, we are eager to determine what nutrients contribute the most to be labeled as unhealthy. Hence, a clustering model will be built to stratify fast food restaurants and their items based on the restaurant’s average nutrition facts and item’s specific nutrition facts respectively. From here, we will visualize the relationship between the prevalence of each cluster and health data of each US state, so that the correlation can be captured. Finally, to add applied features, we will utilize previous results and construct an application to suggest healthy items around users, when given geographical information, such as zip code.

**Introduction**

According to a market analysis[[1]](#footnote-1), globally fast food industry is worth of $570 Billion. While fast food accommodates huge number of busy commuters, travelers, and wage workers who often do not have the time to sit down at a traditional diner or restaurant and wait for their meals, numerous criticisms towards fast food should also not be ignored. For example, negative health effects, alleged animal cruelty, cases of worker exploitation, etc. In this project, we will address one of these criticisms, the negative health effects, and examine its authenticity. We will apply machine-learning techniques to cluster fast food items, verify a correlation between fast food prevalence and health index at state-level and compare fast food’s nutrition fact with recommended nutrition intake from FDA.

**Background**

Fast food consumption in the United States is incredibly large. A research shows that "In the United States, 36.6% of adults consumed fast food on a given day."[[2]](#footnote-2)As an essential part of the American diet people always receive “fast food warning” from everywhere, such as TV shows, newspapers. An article in MedicalNewsToday[[3]](#footnote-3)(2019) suggests that fast food not only has a short-term negative effect, such as promoting more hunger but also has long term negative effects, such as increasing the risk of developing asthma. Every year tons of fast food related topic papers are published; people discuss fast food from various academic fields, such as economics or healthcare. Most of the related work associate fast food with unhealthy lifestyle.

**Motivation**

There exist millions of health problems in the world, but why we are interested in this topic? We love fast food because they usually are cheap and delicious, besides, with the development of fast food chains, people could finish their meals in a convenient and effective way, many busy people even have two meals in fast restaurants a day. Experts stated that fast foods contain excessive fat and oil, which is terrible for humans’ bodies. Are fast foods really unhealthy? Or how can people balance fast food meals so that they can have a healthier meal? Our team is eager to figure out more truths about fast foods.

**High Level problem definition**

According to multiple news reports[[4]](#footnote-4), the United States consumes the most fast food in the world. We consider the data from the United States as the most representative. We collected information including nutrition facts, prices and serving times from over twenty fast food chains, and planned to analyze them with clustering, regression and cross-comparing to obtain solid conclusions about whether consuming fast food can cause negative health effects and develop a reliable recommendation system to suggest people how to eat healthier when dining in fast food restaurants. And hence, we will divide this project into three parts as approaches to our main research question.

**Detailed problem definition and technical challengers**

For obtaining the answer to the question: is the fast food really unhealthy? We explore this problem into three parts. In the first part, we plan to analyze the food composition: whether eating fast food can satisfy daily body demand for keeping healthy? If there is a huge difference existing between fast food nutrition content and the recommended daily nutrition intake content, we can conclude that fast food is not healthy. In the second part, we will approach our main research question by answering a more detailed sub-question: Is there a correlation between fast food prevalence and health index at the state level? To be specific, we want to determine will consume more fast food results in diabetes, or high body weight? If the answer is positive, then we can confirm the negative health effects of fast food. In the third part, we are going to scrape the nutrition facts about the most popular menus from each fast food chain’s official websites, for example, Big Mac Meal at McDonald, Teriyaki Chicken in Panda Express, Spicy Chicken Sandwich in Chick-fil-A etc. We would use the nutrition facts, different meat kinds, prices and even distance from the customers to cluster those popular menus as well as famous fast food chains, labeled them for future use. Then we would recommend customers what should they eat based on their input requirements. There are several technical challenges. Firstly, we need to find out a perfect way to scrape information from the official website and apply proper rules to cluster those menus and restaurants. Secondly, we need to apply Google API to access restaurant locations around customers. Thirdly, we need to make sure that our system can receive customers’ requirements and give out reliable feedback.

**Related Work**

Part 1: As one of the dominated American dietary, there are tons of related studies about fast food. The Obesity Action Coalition posted an article, "Fast Food – Is it the Enemy."[[5]](#footnote-5)(2012) The article listed the calorie content for a fast food meal, which included an entree, a side, and a drink. The results show that people can easily intake over 1500 calories, such high intakes may lead to weight gain or other health conditions like heart disease. However, fast food is the best and only choice for overscheduled and overcommitted American, Thus, fast food at least is partially responsible for the high US obesity rate. In some way, this article uses a straightforward method to analyze the health of fast food. The author only discussed whether fast food is healthy from the perspective of calories. In another paper, Wu, H. and Sturm, R[[6]](#footnote-6) (2012) analyze from the perspective of nutritional content, making the result more convincing. Instead of only considering the fast food, this paper analyzing all types of restaurants. Their study population was the top 400 US chain restaurants based on Restaurants & Institutions magazine's list. They used Logistic regression to predict information accessibility, and statistically analyzed energy and nutrient levels of the menu indicates. Compared to other types of restaurants, fast-food restaurants had significantly higher energy levels on average and not significantly higher levels of fat or sodium. The results show that fast food has too much energy and unbalanced nutrients. This conclusion is that the horizontal comparison of the different restaurant's nutritional content, we cannot get an absolute answer based on this. We can say that "fast food may be less healthy", but we cannot conclude that fast food is absolutely unhealthy. In addition, both Journal “Health Promotion Perspectives”[[7]](#footnote-7)(2015) and “Public Health Nutr”[[8]](#footnote-8)(2015) tried to find whether there is a connection between fast food intake and different diseases. As a result, both of them conclude that Fast food consumption main risk factor for lower diet quality, higher calorie and fat intake, and lower micronutrient density of diet. Most related works release that fast food has adverse health effects.

Part2: Related works in clustering fast food restaurants and detecting the relationship between fast food prevalence and health effects are available. But articles about clustering fast food restaurants are mostly based on geographical information instead of their nutrition facts. For example, Bryn Austin et al.[[9]](#footnote-9)(2011) used a bivariate K function statistical method to quantify the degree of clustering of fast-food restaurants around school-address. A Datafiniti[[10]](#footnote-10) research shows the prevalence of fast food restaurants only but does not correlate with the health index of states. However, we did find an article with a similar topic to our second part. In the article, Maddock[[11]](#footnote-11) (2004) used multiple hierarchical regressions and aggregated state-level data to reveal the relationship between fast food prevalence and obesity.

Part3:[[12]](#footnote-12)Broadly, Recommender Systems can be classified into 3 types: 1. Simple recommenders; 2. Content-based recommenders; 3. Collaborative filtering engines. We will try to implement the content-based recommender. This recommender suggests similar items based on a particular item. This system uses item metadata, such as Locality, Cuisine, rating, etc. for restaurants, to make these recommendations. Additionally, based on a content-based recommender, we will add some features into our system, including features we are going to cluster.[[13]](#footnote-13) The corresponding example code is provided by the source website, and we decided to read, think then modify and finally create our own recommender. For some undefined dishes of some fast food chain (e.g. dishes which miss some important information), followed the instructions on the websites, we would use KNN or K-nearest neighbor to classify them so that customers could receive better recommendations.

Proposed Solution

In the first part, firstly, we need to determine reference data - how much nutrition facts body needs. The daily nutritional goals depend on calorie needs. However, calorie needs vary depending on many factors, including the person's age, gender, and amount of exercise. "2000 calories" is the average calorie needs per day for an adult; we will use this number as our reference calorie intake. Then we choose the nutrition facts that appear on the Nutrition Facts label as the criteria such as protein, carbohydrates, sodium, and sugars to analyze the food nutrition content. For example, as the Nutrition Facts Label shows, 2000 calories from carbohydrates are the recommended intake, to make the results accurate. Secondly, we will collect the fast food nutrition data based on a single restaurant level. We decided to choose the 15 most popular fast-food restaurants in the USA and obtain the nutrition fact content of their top menus. Then compute the average of each nutrition facts criteria content per restaurant. Thirdly, we will compare each restaurant's fast food nutrition data to reference data, to see within the given calorie number (here we use 2000 calories), whether fast food Can provide enough and balanced nutrition.

In the second part, we will first use multi-dimensional clustering models to cluster fast food restaurants based on their average items’ nutrition facts such as protein, sugar, carbohydrates, etc. Ideally, we can generate some different clusters, for example, cluster with high protein but low on sugar, and cluster with high on sugar and calories. From here, we will examine the correlations between each cluster and percentages of diagnosed diabetes and obesity of each state, by first visualizing all states’ diagnosed diabetes, obesity and the prevalence of each cluster’s fast food restaurants by using heatmaps. Ideally, we can spot some compatible trends. For example, we would expect that the visualizations of diagnosed diabetes and obesity highly agree on that of fast food clusters with high on sugar and calories. Then we will perform correlation analysis to validate potential relationships and visualize the results with regression lines.

In the third part, we are not chasing specific solutions since this is more about the application part, but we could still assign a general goal to our recommendation system. We want to develop a system which can receive features including kinds of meat (chicken/beef/pork etc.), fat contents(low/high), calories contents(low/high), protein contents(low/high), price ranges(5~10,10~15,15+ etc.) and locations(customer current locations) then utilize those limits to give out reasonable and reliable recommendations to customers. From our system, customers could have an intuitive impression on how their meals combined and better understanding of how much calories they are going to take. Besides, I think it is a good chance for fast food chains to introduce some health combos to potential customers then hopefully can change stereotypes on fast foods.

1. “Fast Food Industry Analysis 2019-Cost and Trends”, Retrieved from *franchisehelp.com.* Franchise Help. 2018. p. 1. [↑](#footnote-ref-1)
2. “Fast Food Consumption Among Adults in the United States, 2013–2016” Retrieved from *cdc.gov/nchs/data/* [↑](#footnote-ref-2)
3. Huzar,T. “What happens when you eat fast food?”. *Medicalnewstoday,* 2019 [↑](#footnote-ref-3)
4. Sammy, S. Countries That Consume the Most Fast Food or Junk Food. (2013, September 16). Retrieved from *www.therichest.com* [↑](#footnote-ref-4)
5. Muntel, S. (2012, January). Fast Food – Is it the Enemy? Retrieved from *www.obesityaction.org* [↑](#footnote-ref-5)
6. Wu, H. W., & Sturm, R. (2012). Whats on the menu? A review of the energy and nutritional content of US chain restaurant menus. *Public Health Nutrition, 16(1), 87–96. doi: 10.1017* [↑](#footnote-ref-6)
7. Bahadoran, Z., Mirmiran, P., & Azizi, F. (2015). Fast Food Pattern and Cardiometabolic Disorders: A Review of Current Studies. *Health Promotion Perspectives, 5(4), 231–240. doi: 10.15171/hpp.2015.028*  [↑](#footnote-ref-7)
8. Barnes TL, French SA, Mitchell NR, Wolfson J. (2015) Fast-food consumption, diet quality and body weight: cross-sectional and prospective associations in a community sample of working adults. *Public Health Nutrition, 15:1–8. doi: 10.1017/S1368980015001871* [↑](#footnote-ref-8)
9. [S. Bryn Austin](https://ajph.aphapublications.org/author/Austin%2C+S+Bryn)ScD, Steven J. MellyMS, Brisa N. SanchezScM, Aarti PatelBA, Stephen BukaScD, and ,[Steven L. Gortmaker](https://ajph.aphapublications.org/author/Gortmaker%2C+Steven+L)PhD (2011), Clustering of Fast-Food Restaurants Around Schools: A Novel Application of Spatial Statistics to the Study of Food Environments [↑](#footnote-ref-9)
10. The Top Cities and States for Fast Food Restaurants. (2018) Retrieved from *datafiniti.co* [↑](#footnote-ref-10)
11. Maddock, J. (2004), The Relationship between Obesity and the Prevalence of Fast Food Restaurants: State-Level Analysis, *American Journal of Promotion Health* [↑](#footnote-ref-11)
12. Chauhan, N(2019) How To Build a Restaurant Recommendation Engine (part-1) Retrieved from *https://medium.com/analytics-vidhya* [↑](#footnote-ref-12)
13. Chauhan, N(2019) How To Build a Restaurant Recommendation Engine (part-2) = Retrieved from *https://medium.com/analytics-vidhya* [↑](#footnote-ref-13)