**Data Report on Fast Food Analysis**

Samantha Paige Minchew

Professor Satvika Marrapu

Data Analysis and Knowledge Discovery, INFO 4670, Section 001

Midterm Assessment

14 October 2023

This is a report for the class Data Analysis and Knowledge Discovery taught by Professor Satvika Marrapu. In this report, I will closely examine fast food data to create an in-depth analysis that contains a comparison analogy between various fast-food chains. This report includes data model techniques that gathers information with pivot chart and tables, slicers, pareto charts, histograms, lookup and match functions to find which fast-food restaurants have healthier eating options, and which food items have higher caloric and unhealthy dietary alternatives.

**Data**

This report is listed as continuous data rather than discrete data because menu items are constantly being added and recipes are often changed. The data set is not normalized since it has information from various food items and is collected from different restaurants. There is a plethora of fast-food chains, and countess menu items that can make it challenging to collect intuitive data on. In this analysis, data is collected in main fast-food chains such as, McDonald’s, Chick-Fil-a, Burger King, Arby’s, Subway, Dairy Queen, and Taco Bell. The data was taken from Kaggle but was originally posted on “fastfoodnutrition.org.” The data set is structured by visualizing macronutrients from different food items. The focus of this report is to compare and analyze which menu items have more calories, protein, sodium, etc. to be able to make conscious eating decisions. The time interval for this data is collected during current time so it is still relevant information that can be used to make health-conscious decisions. Originally, I had used Mean, Median, and Mode, and standard deviations; however, these functions did not gather the information needed and was irrelevant to my data focus. I shifted the report to focus on information that would be useful to the thesis of this report. Since the data was so extensive, useful information was gathered by focusing on charts and SUM functions to visually analyze the purpose of this report. Data is collected manually via excel spreadsheet, and more automatically using rapid miner and tableau. Even though tableau is not required for this class, I chose to include it in this report because it was able to hold more information and create visuals I wanted.

“Fast Food Calories Data (Kaggle 1),” lists this data set in detail by listing over 500 unique values and cross examining them. In their data set they explore mean and standard deviation, but only as a total factor and not how it relates to other data. Kaggle referenced “fastfoodnutrition.org,” which gives a deeper look into the data that was collected from various restaurants. This website lists nutrition facts from every fast-food chain, mainstream coffee shops, and nice restaurants.

**Objectives**

Using this data set, I will be able to analyze and cross examine information from various fast-food chains and take an in-depth investigation of nutritious facts. With this report I will be able to answer the following questions:

1. Which fast-food restaurant has higher protein food, and which has higher calorie food?
2. Which menu items have more protein, and which have more calories and sodium?
3. Does calorie and protein food items co-exist (if so, that would not make the higher protein food items any more nutritious than lower protein food)?
4. Without specific menu items, which food chain has better nutrition?
5. Are there healthy diet options in food chains or are they higher in calories and empty nutritious values?

**Methods**

* 1. ***Tools***

In this data report, excel, rapid miner, and tableau are used. These tools are great for analyzing and extracting data for a comparison and contrast output. I was having issues with using excel because it kept shutting down and working very slow, this was because of the large data file. I used excel to complete the requirement but also added tableau for the automatic graphs it was able to create regardless of the file size.

**Data pre-processing**

* Pre-processed data, I first transferred the graph from the original file to an excel spreadsheet.
* There were over 500 different variations of data, so it was overwhelming to work with.
* I uploaded to data to rapid miner and tableau to get an easy visual of the information rather than going through every single set.
* After getting an understanding if the data I was dealing with, I went back to excel to use the mean, median, and mode functions and calculated the standard deviation.
* After those calculation, I re-examined the data set because those functions did not have the data that I originally wanted to find.
* I set aside those functions and focused on creating pivot charts and histograms from different restaurants and comparing them to each other.
* I then manually calculated the SUM before creating a graph that compared the SUM of nutrients from all food items at all the restaurants.
* I added match functions to be able to find any information I needed without having to sort through all the data.
* I also added lookup functions to be able to compare calories vs. protein since that was my original intended focus for this report.

**Data Analysis**

* The model used in this report is pivot tables and charts, histograms, pareto charts and adding slicers to be able to change the data according to my questions.
* I also used HLOOKUP and VLOOKUP functions and MATCH functions.
* I chose these methods so that I was able to find the answers needed and create a visual graph to demonstrate the data without having to sort through every set.
* I found that graphs and slicers allowed the information to be processed the quickest since there was so many nutrient items and various menu categories.

**Results**

-Using visual analysis methods, I was able to retrieve the information that I wanted to answer the questions I had about my data set.

-I was able to see nutritious facts and understand how much macronutrients were used.

-After reviewing the graphs, I was surprised to see how much sodium was in the ingredients, I was expecting it to have more calories and higher protein content.

-I was also surprised about how little nutrition value was in foods that were advertised as “high protein” meals and how much sodium and calories were in almost every food item

***Figure 1.A: Rapid Miner***

**A screenshot of a graph

Description automatically generated**

Figure 1.A demonstrates the pareto line of calories between all food items on rapid miner. Already you can see that the lowest caloric intake is 380 and goes up to 2430. However, most food items lie within 380-710 range.

***Figure 1.B: Rapid Miner***

**A screenshot of a graph

Description automatically generated**

This graph uses the same data to examine the amount of protein in food on rapid miner. The highest amount of protein is 64, and the lowest being 5 grams. Most food items have a protein range from 23 to 13.

***Figure 2.A: Tableau***

A screenshot of a computer

Description automatically generated

Arranging all food items in an order from most calories to least, it’s easy to see that protein content and calories are somewhat connected. The more calories, the more protein (not consistent enough however, to make an informative decision on). Although the highest protein content has more calories, a lot of high calorie food items don’t have any protein at all.

***Figure 2.B: Tableau***

A screenshot of a computer

Description automatically generated

This graph in Tableau shows the amount of sodium in comparison to cholesterol, calories, and protein. You can see that sodium levels stay the same but are somewhat inverted to cholesterol levels. Food with higher sodium seems to have somewhat less cholesterol amounts, whereas protein and calories go hand in hand to each other.

***Figure 2.C: Tableau***

A screenshot of a computer

Description automatically generated

This graph compares which food chain has more products with a higher protein count. Surprisingly, subway had higher calorie food items but not the highest calories overall. Taco Bell had the most calories in their food items without much protein items. I was surprised to see Chick-Fil-A was at the bottom of both comparative lists.

***Figure 2.D: Tableau***

A screenshot of a computer

Description automatically generated

Again, Taco Bell has the highest caloric menu item that is not in a direct relation to the amount of protein. Thus, Taco Bell would not be a good choice for high protein and lower calorie and sodium level food items.

***Figure 2.E: Tableau***

A screenshot of a computer

Description automatically generated

This graph demonstrates the amount of SUM calories between different food chains. It also compares the SUM of protein, sodium, cholesterol, and calcium for a more in depth and interactive graph.

***Figure 2.F: Tableau***

A screenshot of a computer

Description automatically generated

I personally really enjoyed using tableau for this assignment because it visualized very specific data to be able to make an informative decision about food items. This graph goes through different restaurants, analyzing and comparing the number of macronutrients in each menu item.

***Figure 3.A: Excel***

A screenshot of a computer

Description automatically generated

On excel, working with SUM was found helpful. Since there was such a big data set, using excel to visualize and slice information helped create the answers I had. Using slicers helped me create an interactive chart that listed which restaurant, and which macronutrient items I had questions on. Using this tool, I was able to select not only where to eat, but which food items were healthier.

***Figure 3.B: Excel***

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Using pareto charts on each restaurant as well as creating histograms helped analyze calories vs. protein intake more in depth across each fast-food chain. This was very interesting to me because I received answers that I wasn’t expecting to have. I was shocked to see the extremely high levels of sodium that were in almost every restaurant, and I was also able to see that calories and sodium levels almost copied each other on the histogram.

***Figure 3.C: Excel***

A screenshot of a computer

Description automatically generated

Lastly, on my report I used LOOKUP functions and MATCH functions to help identify food items quickly and report the macronutrients listed. This is helpful because if I wanted to know what was in a menu item, I could use MATCH function to find it, and LOOKUP functions to examine nutritious value of the item.

**Discussion**

After extracting the information from the data set, I was surprised by the conclusion I had. There was extremely high levels of sodium and cholesterol. At the beginning of this report, I was expecting to see just a high number of calories. I was expecting calorie levels to be the biggest issue with fast-food menu items, but after this report I saw that it was also the sodium levels. Rather than analyzing calorie intake, it is important to also find out what the sodium levels are. I was also surprised to see how little of protein were in food items advertised as “healthier” or “high protein.” This report really changed how I view eating at fast-food restaurants.

The limitations I had in my report include being able to correctly compare different restaurants from each other since some had more items listed than others. It made it hard to look at the total SUM of each place and cross-analyze it to another one if there wasn’t an equal amount of food items to compare. This is why I created the histogram charts and focused more on graph representation over numbered reports. If I were to do this assignment again, I would rather focus on macronutrients from a single restaurant rather than multiple so I would be able to analyze in more detail the menu items. I found that there was a limitation to a lot of my calculations because there were so many restaurants that it made it difficult to really compare food items to each other. I really wanted to use mean, median, mode and standard deviation techniques; however, it did not make sense with my report, and I could not find a place for the calculation to fit any of my questions or answers.

**Evaluation and Conclusion**

During this assignment I learned how to use excel for my own benefit by finding answers to data set questions I had. I also learned thar even though excel is very detailed, it takes a lot of time to manually create the representation I wanted. Being said, I really enjoyed using tableau for this project because it was a lot quicker to retrieve answers to the questions, I had about my data set. I found excel to be very slow, did not load properly, and kept deleting my files after I had too much data on it, which made it feel not very dependable. I had to keep saving my progress and closing out the spreadsheet to finish the assignment. I was also disappointed how rapid miner responded to my data set. I am not sure if it was too much information or if it would not categorize it correctly, but it was difficult for me to retrieve data on it like I expected it to. It kept comparing the data as a whole and would not organize it like I tried to. Overall, this assignment gave me the technical skills to be able to analyze any data set and retrieve answers and questions to any data set. Now I have learned how to retrieve data that I need and be able to use it for my personal use. If I was going to eat at a fast-food restaurant and be health conscious, I could decide which place to eat at using SUM function and know which food item to get according to what macronutrients I was looking for.

**References**

*Fast Food Restaurants Nutrition Facts*. FastFoodNutrition.org. (n.d.). https://fastfoodnutrition.org/fast-food-restaurants

Kapadnis, S. (2023, October 13). *Fast Food Calories Data*. Kaggle. https://www.kaggle.com/datasets/sujaykapadnis/fast-food-calories-data/