
SIMON FRASER
UNIVERSITY

IAT 814:Visualization and Analytics
Instructor: Lyn Bartram
Spring 2021

# Fast Food Analysis Notebook

Members:
**Kenneth Lau**
**Srijeev Sakar**

# 1. OVERVIEW OF PROJECT

The fast-food analysis project facilitates interactive visualizations in a top-down manner to help nutritionists, dieticians and analysts develop a better intuition of what popular chains in the United States such as McDonald's, Burger King, Chick-fil-A, Subway, Starbucks have to offer.

The project looks at fast-food categorization pricing, compares their nutritional compositions, and sees how distributed they are across each state. Further, we account for factors like obesity, income, population and plot their relationships with how many fast food options there are.

The project notebook consists of sections based on the purpose and utility of the visualizations. Information on how to interact and use each visualization is available in the visualization design section of this report.

The first section showcases all the menu items offered by the five fast-food chains. It provides a medium to explore the relationship between fast-food prices and nutrition. We have also included an option to compare these menu items by particular nutritional metrics.

The second section allows the user to compare multiple sandwiches and drinks to help them decide what a nutritionally better choice is. It can help consumers make better decisions. For example, if a consumer is interested in a low sugar drink or a high protein sandwich, it helps to know what other relevant options are available.

The third and fourth section explores the relationship between income/obesity/population in each state and the number of fast-food choices available. It also accounts for the correlation between these factors head to head. It may be tough to conclude relationships, such as a direct relationship between income and the number of fast-food choices, obesity and population, etc. However, we believe that leveraging this data through interactive visualizations will help users discover trends and patterns for analysis.

The fifth and final section summarizes all the available menu items and their prices using an easy-clicking interface.

The report is divided into sections with an appendix in a fashion that's similar to our project. We believe it helps guide new users/target audiences in a step-by-step manner.

# 2. INTRODUCTION

## a. Scope

Over a third of all restaurants in the United States sell fast food. There are approximately 250,000 restaurants that represent 50,000 chains[1,2]. Fast food has become a way of life. They are cheap, tasty, and convenient. Most of these foods are high in fats and carbs and lack fibre and vitamins. However, most of the offerings by fast-food restaurants are not necessarily poor choices. The key is to find a balance in each meal, and we hope to address that via our work. The project aims to help our targeted audience understand the menu distribution, pricing, and nutritional value of fast food options from popular brands. We also take it a step further by analyzing the relationship between factors like income/obesity/population per state and how they compare to the number of fast-food options available.

## b. Users/Audience

The targeted audience consists of nutritionists, marketing analysts, researchers, and anyone curious about the intricacies of fast-food offerings.

Nutritionists cater to their client's needs. If a client is addicted to fast food, it may be wise to explore programs that don't rule out all fast food choices. The analysis notebook can help nutritionists locate healthier categories like salads in different restaurants. It can also facilitate comparisons between those options.

Marketing analysts can use the population/obesity/income information alongside fast-food data to target different niches of people. For instance, California has a large population, moderately high incomes, and lower levels of obesity. It could be an opportunity to advertise Subway's healthier subs and salads, as California has the most number of Subways[3].

Health-conscious and nutritionally curious people can leverage our product to gain better insights on what food items to eat to satisfy their nutritional requirements, in turn, improving food habits and awareness.

# 3. PROBLEM/DOMAIN QUESTIONS

1. How do fast food items from highly coveted chains compare in terms of their nutritional composition?
2. Is there a relationship between nutrition and expense?
3. How does income/obesity/population relate to each other?
4. How does the distribution of fast-food restaurants differ in US states, and is there a relation to markers such as income/obesity/population?
5. How do fast-food sandwiches and drinks compare in terms of nutritional value?
6. For each fast-food restaurant, what is the division in product categorization and pricing?

# 4. DATA DESCRIPTION

## a. Sources

There were a plethora of datasets used for this project that were gleaned and manipulated from various sources. The datasets consist of information about fast food items, geographic information for restaurant chains, and income, obesity, population data for each US state. The sources include Kaggle, Deptofnumbers, America's health rankings, Macrotrends, and the official fast food menu prices website. We cited them in this report with reference links.

## b. Dimensions

- Geographic information for fast food restaurants

| Column Names | State | Number of restaurants in state |
|---|---|---|
| Data Types | String | Integer |

Date Source: Kaggle [4]

- Fast-food menu items with nutritional compositions

| Column Names | Restaurant | Category | Items | Cost | Calories | Protein | Fat | Carbs | Sugar | Sodium |
|---|---|---|---|---|---|---|---|---|---|---|
| Data Types | String | String | String | Float | Integer | Integer | Integer | Integer | Integer | Integer |

Date sources: Kaggle [5-9] Fast Food Prices [10]

- Geographic information - income per state

| Column Names | State | Income |
|---|---|---|
| Data Types | String | Integer |

Date Source: Deptofnumbers [11]

- Geographic information - obesity per state

| Column Names | State | Obesity |
|---|---|---|
| Data Types | String | Integer |

Date Source: America’s health rankings [12]

- Geographic information - population per state

| Column Names | State | Population |
|---|---|---|
| Data Types | String | Integer |

Data Source: macrotrends[13]

- Historic information for time series data - income/population/obesity level per state

| Column Names | State | Indicator | Year | Value |
|---|---|---|---|---|
| Data Types | String | String | Integer | Integer |

Data Source: Deptofnumbers [11], America's health rankings[12], macrotrends[13]

## c. Data Cleaning/Manipulation

Data munging was an integral part of our project before visualization development. The most crucial steps in terms of transforming the data to a suitable form for our analysis included:

1. Merge various fast-food restaurant datasets into a larger dataset for ease of access.
2. Combine pricing data from a reputable source with the larger dataset.
3. Locate and remove redundant food items such as a Big Mac and a Big Mac without pickles.
4. Rename the naming convention of fast-food items by adding labels. For example, a cheeseburger from Burger King gets denoted with BK, and a cheeseburger from Mcdonalds gets McD. It helps the code recognize these two items as separate entities.
5. Many items were missing prices as well as categories such as breakfast, drink, etc. We manually located them and assigned labels.
6. Appended state codes to the datasets using a Python dictionary to facilitate plotting on a map.
7. Starbucks has an extensive drink menu with redundant drinks showing up with different minor compositions and cup sizes. We standardized it to medium and removed repetitive products.
8. We filtered out drinks and sandwiches for a comparison-based analysis.
9. The time-series data from the year 2000 to 2020 was challenging to collect. We used a simple CSV format and plugged in the data from various sources for each state.
10. We standardized the data collection operations in our code by importing all the datasets from our public Github repository, allowing anyone to access our finalized data sources.

# 5. VISUALIZATION DESIGN

## a. How to run the code

1. Unzip the provided zip file or access the public Github repository https://github.com/srijeev-ds/IAT
2. Use either the fast_food_analysis_IAT_project.ipynb notebook or the fast_food_analysis_IAT_project.py file in the 'code folder'.
3. As long as you have dash installed (pip command included), you may choose to run the .ipynb file on jupyter notebooks or the python file locally.
4. The project video and report contain instructions and details on how to explore the visualizations.

**Alternatively, you can visit our deployed dashboard via the link:**
**https://fastfood-analysis-app.herokuapp.com**

## b. Visualization Designs

### Visualization 1

**Description of visualization:**

- The set of visualizations above consists of bar charts placed side by side. There are two sets of drop-downs above them.
- The first bar chart (left) displays the nutritional composition of the selected item.
- The second bar chart (right) displays the distribution of selected fast food menu items.
- The first drop-down allows a user to select one or more restaurants to update the bar chart displaying menu items.
- The second drop-down updates the y-axis upon selection of cost/nutrient. It facilitates a comparison amongst fast-food items.

**Justification of design choice:**

- The goal of these interactive visualizations is to represent the values of price/nutritional components.
- The use of bar charts with standardized visual encoding colours (for restaurants and nutrient values) helps the user compare menu items and further drill down on them.
- We had initially used pie charts on the left to represent nutritional composition. But it didn't make sense to showcase a proportion.
- A bar chart with a standardized unit of 'grams' facilitates easier comparison.
- Further, the use of clear drop-down menus, options, and labels further strengthen the design choice.

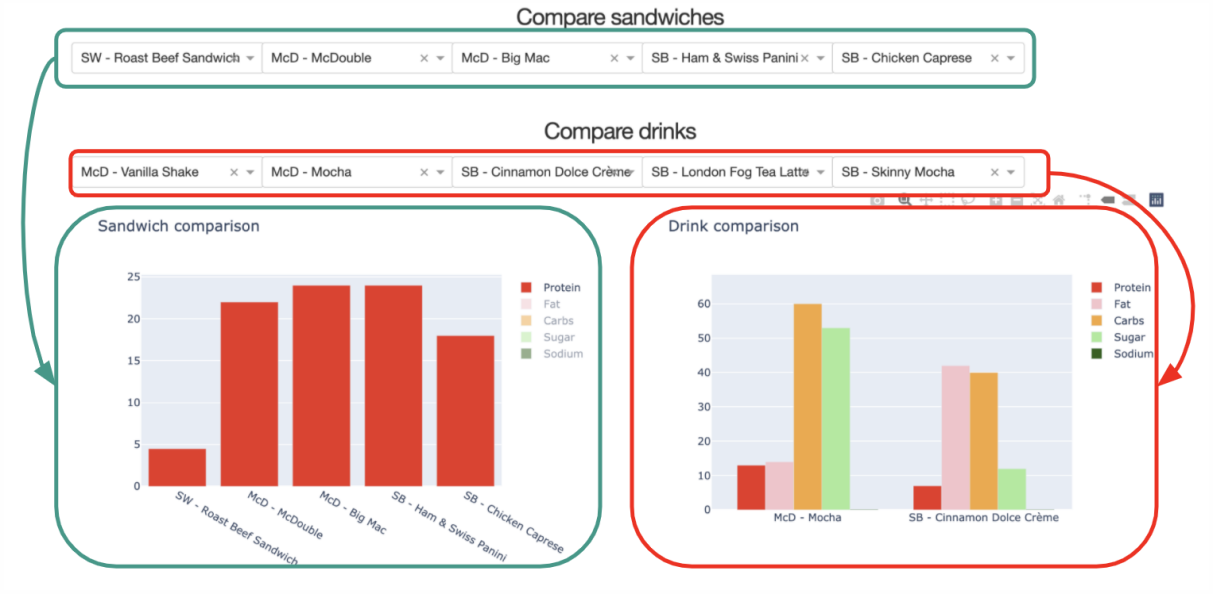
**How it relates to our problem questions:**

- The bar chart on the right side displays menu item names on the x-axis and cost on the y-axis by default. It helps any user spot items that are expensive. If the user clicks on an item, a nutritional composition, the left bar chart shows its nutritional breakdown. By exploring the pricing vs nutrition of these options, a user can analyze the relationship between pricing and nutrition values.
- The user can change the comparison metric from cost to protein/carb/sugar/fat/sodium using the second drop-drown. It allows comparisons based on nutritional metrics.

**How do they work together:**

- The user can click on any item on the menu distribution bar chart (right) to update the nutritional breakdown bar chart for an item (left).
- Items may be removed or added using the first drop-down, and comparison metrics can be changed using the second drop-down.
- The user can also hover and select a subsection of the displayed graph to zoom in. It enables an easy drill down.

# Visualization 2




## Description of visualization:

- The two sets of visualizations consist of grouped bar chart visualizations placed side by side. Two rows of five drop-downs are above them.
- The first row of drop-downs allows the user to select up to five sandwiches from any restaurant.
- The second row of drop-downs allows the user to select up to five drinks from any restaurant.
- The labelled grouped bar charts facilitate the comparisons of drinks and sandwiches.
- The legends represent nutritional components and facilitate further drill-downs.

## Justification of design choice:

- The goal of these grouped bar chart visualizations is to represent the nutritional components of the menu items selected by the user for comparison.
- Grouped bar charts make it easier to compare our items as it places them on a common x-axis.
- These bar charts make use of the uniform visual encoding colours selected for nutrients.
- Further, clicking on the legend can allow comparison based on one or more nutrients.

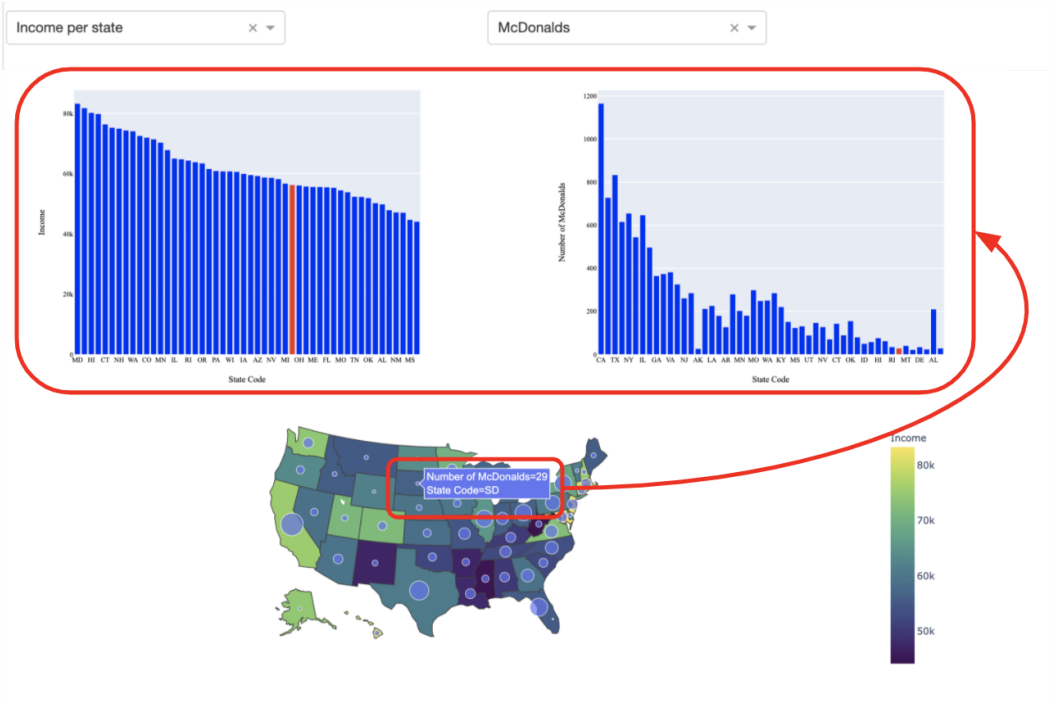
## How it relates to our problem questions:

- These two sets of visualizations build on the same questions mentioned for Visualization 1.
- It helps us see how fast-food sandwiches and beverages from any restaurant compare in terms of nutrition.

**How do they work together:**

- Once the user selects sandwich/beverage options using the provided drop-down, the group bar chart updates its values with the nutrient composition of the selected items.
- The user can click on any of the displayed legends like protein, carbs, etc., to compare by fewer metrics. The example figure compares sandwiches using protein. The other metrics have been de-selected through clicking.
- The user may also hover and select a subsection of sandwiches or drinks to zoom in for a closer comparison.

# Visualization 3




**Description of visualization:**

- The interactive visualization consists of two drop-down options that update all three visualizations.
- These visualizations are connected and update values simultaneously dependent on the requirement.
- The first drop-down consists of income/obesity/population per state.
- The second drop-down consists of how many Burger Kings, McDonald's, Subways, Chick-Fil-A's, and Starbucks each state has.
- The first bar chart represents the choice from the first drop-down on the y-axis vs the US states on the x-axis. For example, Obesity in each US state.
- The second bar chart represents the choice from the second drop-down on the y-axis vs the US states on the x-axis. For example, the number of Chick-Fil-A's in each US state.
- The choropleth map represents three variables. The choice from the first drop-down colour encodes the map. The value from the second drop-down is aggregated and

binned in each state. State codes decide upon the geographical boundaries as we set the map location to the United States. For example, the colour represents the population of each US state, and the bins represent the number of Subways.

**Justification of design choice:**

- The goal is to see how income/population/obesity varies in each US state. We would also like to know the number of fast-food options in each state.
- Bar charts are an ideal way to represent these ideas.
- The left bar chart helps us analyze and compare how income/population/obesity varies in each US state.
- The bar chart on the right represents the number of selected available fast-food options in each state.
- The choropleth map is a way to combine the data represented by the two bar charts into a singular visualization (three variables).
- The boundaries on the map represent the states. The income/population/obesity values are encoded using colour, and the number of fast-food restaurants based on our selection gets binned in each state.

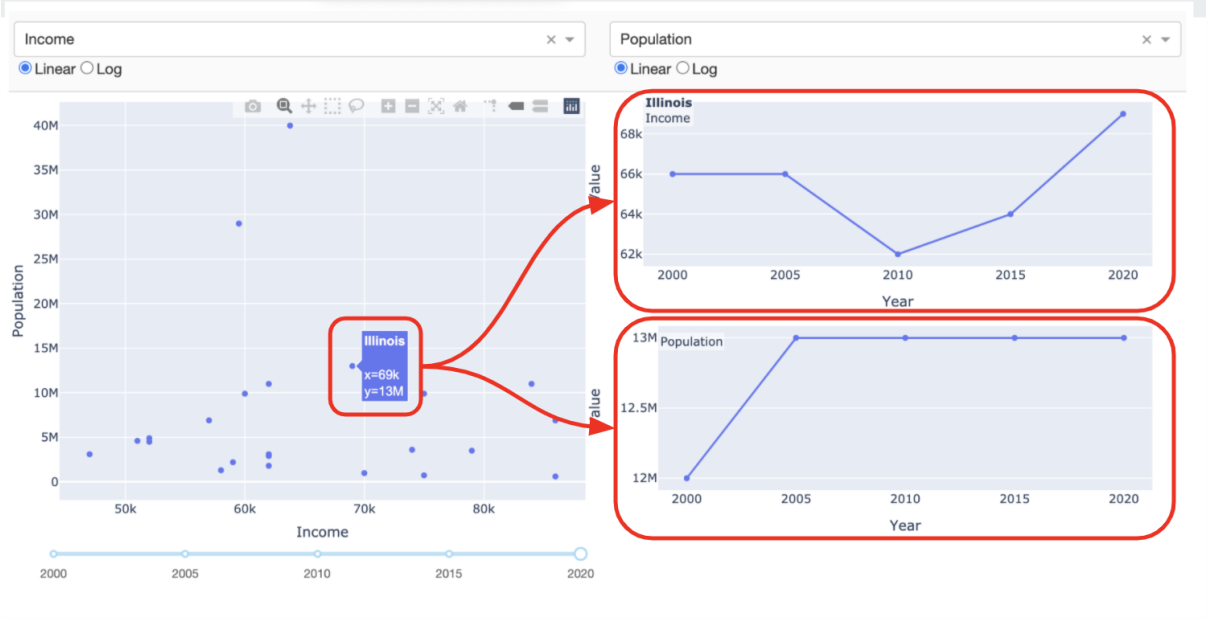
**How it relates to our problem questions:**

- The visualizations help explore the relationship between the number of available fast-food options and markers such as income/obesity/population.
- The bar charts and the choropleth map help capture the state-to-state difference for these metrics.

**How do they work together:**

- The first drop-down updates the left bar chart and colour codes the choropleth map based on the selected income/obesity/population marker.
- The second drop-down updates the bar chart on the right and bin values in each state of the choropleth.
- Hovering over any bar in the bar charts or US states/bins in the choropleth map updates the tooltip.
- The map interacts with the two bar charts by highlighting appropriate data in red based on the US state the user hovers over.

# Visualization 4




## Description of visualization:

- The interactive time-series visualization consists of two drop-down menus and a slider to update all three visualizations.
- The three visuals (scatter plot and two-line charts) are connected. The values change based on the hover/zoom/slider/option selected.
- Both the drop-downs consist of the options income, obesity, and population. It enables insights beyond fast food over a long window of time. (2000 to 2020)
- The large scatter plot compares the selected options from the drop-down menu against each other. It can make comparisons on a linear or log scale, depending on how skewed the values are. For example, Population vs Income for each US state.
- The points on the scatter plot represent US states. Hovering over any point updates the two line charts on the right side. For example, if the scatter plot represents Population vs Income, and the user hovers on Illinois, the line charts update with income and population change data for Illinois over the years.

## Justification of design choice:

- The goal is to see how two variables (for example - obesity and population) varied over time in each US state to facilitate comparisons and drill-downs.
- The time-series scatter plot represents US states through points. The selected metrics from the drop-down update the x and y axes. The default value for the slider is 2020.
- The line charts update based on the point (US State) hovered over in the scatter plot. The slider updates the values for each US state based on the year.
- A radio button allows the user to make comparisons using either a linear or logarithmic scale. It helps handle skewed values.

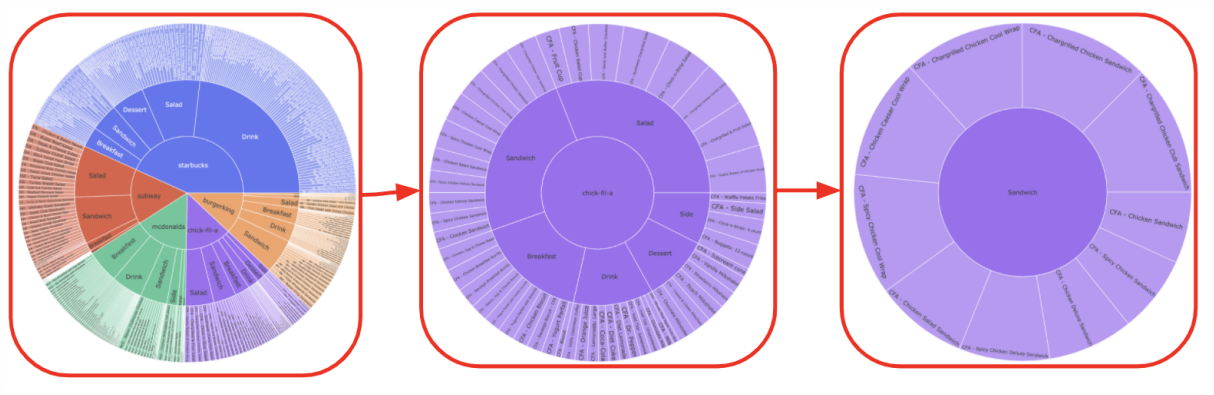
**How it relates to our problem questions:**

- This set of visualizations help a user determine how variables such as income/obesity/population relate to each other over time.
- The time slider and interactive line charts help illustrate the relationship further when a user hovers over a data point.

**How do they work together:**

- The two drop-downs update the axes of the scatter plot (example - Income and Population). The time slider sets the year for the data (example - 2015).
- When a user hovers over a data point in the scatter plot (example -Illinois), the two interactive line charts update with the selected metric value over the years. (example - Income from 2000 to 2020, Population from 2000 to 2020).
- The radio button can change the scale of comparison from linear to log.

# Visualization 5




**Description of visualization:**

- The hierarchical pie chart visualization displays all available fast-food items categorized into breakfast food, desserts, etc.
- The restaurants are colour-coded. A user can gain more information by clicking on a restaurant to drill down.

**Justification of design choice:**

- We wanted to display all our restaurants' food categories, menu items, and prices to provide users with a one-stop visualization in the notebook for that kind of information. We also wanted to summarize all this information in a single visualization.
- The colour encoded interactive hierarchical pie chart is a visualization that captures this purpose.

**How it relates to our problem questions:**

- This set of visualizations help a user determine how variables such as income/obesity/population relate to each other over time.
- The time slider and interactive line charts help illustrate the relationship further when a user hovers over a data point.

**How do they work together:**

- The user can click on any restaurant to enable a drill-down. For example) The user clicks on Chick-Fil-A to update the hierarchical pie chart with categories such as sandwiches, salads, desserts, etc. The user then clicks on sandwiches to display all available products with their prices.
- The visualization enables an easy method to capture menu distribution, summarize products, categorize them, and capture prices.

# 6. FURTHER WORK

## a. User Testing

The product was subject to user testing during our entire development process. We believe in looping in feedback during the development to promote a higher quality work vision.

In terms of the timeline, we received feedback from multiple users to improve our project. The identities of the users apart from the TA and Professor have been kept anonymous for this report.

**PMP Big Data Student 1:** The student suggested comparing fast-food sandwiches and drinks head to head.

We implemented this idea using our set of grouped bar chart visualizations.

**PMP Big Data Student 2:** The student mentioned that the hierarchical pie chart took too much space, especially as the first visualization.

We decided to use it as a summary tool for our menu items and moved it to the bottom.

**PMP Visual Computing Student 1:** The suggestion was to incorporate bar charts above the map to see the distribution of income/population/obesity and the fast-food distribution in each state, without referring to the choropleth map.

We added these interactive bar charts right above the choropleth map.

**TA - Maryam Rezaie:** To add more interactivity, such as make the map interact with other elements, and compare the variables income/obesity/population against each other. She also suggested better use of spacing.

We then made the map interactive with the presented bar charts through highlighting + hovering. We also decreased the size of our grouped bar charts and placed them side by side through our work on a CSS stylesheet.

**Professor Lyn Bartram:** To focus on colour encoding, work on the placement of visuals, and justify the use of our pie chart visualization.

We used a standardized colour encoding for all our visualizations representing nutritional information. We replaced the pie chart showcasing meaningless proportions with a bar chart that displays the values of the nutrients in grams. The sandwich and drink comparison visualizations were moved up to the top, as they were analogous to the first visualization.

## b. Improvements

**In terms of questions:** Nutrition and fast-food analysis are broad topics. We are overall satisfied with the questions we addressed. If we had more time and data, we would address other questions focused on users like:

1. What are the types of choices available to patients with diabetes, heart conditions, etc.?
2. If a user inputs a location, can we locate the number of fast-food places near that location and filter it by healthier choices/user preferences?

**In terms of visualizations:** We tried to combine visualization 1 (bar chart + bar chart) and visualization 2 (grouped bar chart) through clicking and adding items.

> The idea was to click on items on the distributed bar chart from visualization 1 to update the comparison group bar charts from visualization 2. There were multiple issues with the hover and click code, and after two days of trying, we didn't have sufficient time to pull through with that idea.

**In terms of what didn't work:** We initially implemented a pie chart as a design choice to display nutritional composition. The professor pointed out that it isn't necessarily insightful as a pie chart represents proportions. We hence opted for a bar chart to show the distribution of the products in grams.

**If we build version 2:** We would think of creative ways to connect all our visualizations to prepare a highly interactive dashboard. We would also address some of the questions stated above. We would also build a horizontal one-stop approach for the second version of our dashboard compared to our current vertical top-down scroll approach.

# 7. CONCLUSION

Nutrition and fast food have become popular subjects for discussion within our community. With an ever-increasing obesity rate, a slow rise in income, and a busy lifestyle, fast food has become a go-to choice for most of us. Most of us consider fast-food offerings unhealthy but prefer them over healthier choices because of convenience and lower prices. We believe that the key to a healthier lifestyle is balance. There are fast-food restaurants available that offer healthy alternatives. We must be aware and recognize that.

Throughout this project, we had the opportunity to develop and design thought-provoking interactive visualizations to understand fast food. The visualizations helped categorize items, explore prices, and compare popular items through several metrics. Our visualizations also helped correlate but not conclude relationships between obesity/population/income and the distribution of fast-food restaurants. We also looked into how these metrics compared against each other for each US state. We hope our notebook has helped provide nutritionists/analysts/marketers better insights on what's available and has helped users pick meals with greater intention. Nutrition has tremendous scope. In terms of our future works, there is a lot to be explored.

We thank Professor Lyn Bartram, TA Maryam Rezaie, and our supportive classmates for the constant feedback and support throughout our project development process.

# 8. Citation

1. https://www.statista.com/statistics/196619/total-number-of-fast-food-restaurants-in-the-us-since-2002/#:~:text=How%20many%20fast%20food%20restaurants,thousand%20over%20the%20past%20decade.
2. https://hubworks.com/en/blog/statistics-about-americas-biggest-fast-food-chains.html#:~:text=If%20you%20are%20wondering%20how,50%2C000%20scattered%20across%20the%20nation.
3. https://www.scrapehero.com/location-reports/Subway-USA/
4. https://www.kaggle.com/datafiniti/fast-food-restaurants
5. https://www.kaggle.com/davinm/subway-restaurant-nutrition-data/version/1
6. https://www.kaggle.com/mcdonalds/nutrition-facts?select=menu.csv
7. https://www.kaggle.com/starbucks/starbucks-menu?select=starbucks-menu-nutrition-drinks.csv
8. https://www.kaggle.com/starbucks/starbucks-menu?select=starbucks-menu-nutrition-food.csv
9. https://www.kaggle.com/freeno1d/chickfila-nutrition-fact-anly501
10. https://www.fastfoodmenuprices.com
11. https://www.deptofnumbers.com/income/
12. https://www.americashealthrankings.org/explore/annual/measure/obesity/state/ALL
13. https://www.macrotrends.net/states/

# 8. INDEX