Capstone Project

Name: Sakthieaswaran C
Bootcamp: Data Analyst Bootcamp
Workshop ID: DAT-V862-12023266/DA BATCH AUGUST
email: sakthi.easwar01@gmail.com
Mobile: 9597493609
Project Title : NutriCal - McDonald's Menu Nutritional Analysis

Summary

As a data analyst tasked with delving into the nutritional content of McDonald's menu items, my objective is to conduct a comprehensive examination of the diverse array of offerings available at McDonald's outlets worldwide. This analytical endeavor is designed to yield valuable insights into the calorie count and nutritional composition of the various menu items.

McDonald's, a ubiquitous global fast-food chain, is renowned for its extensive menu that caters to a broad spectrum of tastes and preferences. The task at hand involves a meticulous examination of the nutritional content of these menu items, delving into not only the calorie count but also a detailed analysis of the broader nutritional spectrum.

To initiate this analysis, it is imperative to consider the variety of items on McDonald's menu, ranging from classic burgers and fries to salads and beverages. By breaking down the components of these items, we can unravel the intricate web of calories, fats, proteins, carbohydrates, and other essential nutrients that contribute to the overall nutritional profile .This examination would encompass not only the caloric content but also potential implications for sugar intake and other additives.

Data Analysis Approach and Methodology:

Technical Aspects:

Language Used : Python is a popular programming language widely used in the field of data analysis. Its simplicity and versatility make it an excellent choice for handling and manipulating data .Python's ecosystem, combined with its readability and ease of learning, has contributed to its dominance in data analysis and has made it a preferred language for data scientists and analysts.

Libraries Used : **Pandas** is a powerful library for data manipulation and analysis. It provides data structures like DataFrames and Series, making it easy to handle and analyze structured data.

Matplotlib is a widely-used plotting library for creating static, interactive, and animated visualizations in Python.

Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Plotly Express: Plotly Express is a high-level interface for creating interactive visualizations. It is built on top of the Plotly library and simplifies the creation of complex plots

Data Analysis Approach and Methodology:

- ✓ **Module imports :** Importing necessary imports to perform all activities for the analysis
- ✓ **Data Collection:** Acquiring McDonald's menu data containing nutritional details from a chosen source.
- ✓ **Loading the Data:** Using pandas to read and load the dataset into a structured format for analysis.
- ✓ **Data Preprocessing:** Cleaning and organizing the dataset, addressing missing values and converting data types.
- ✓ **Inspecting the Data**: Examining the dataset's structure and initial content to identify patterns or outliers.
- ✓ **Finding Null Values:** Identifying and handling null values to ensure data completeness.
- ✓ **Finding Duplications**: Detecting and managing duplicate entries to maintain data integrity**.**
- ✓ **Getting Descriptive Statistics:** Calculating descriptive statistics to gain insights into the numerical aspects of the dataset.

Data Analysis Approach and Methodology:

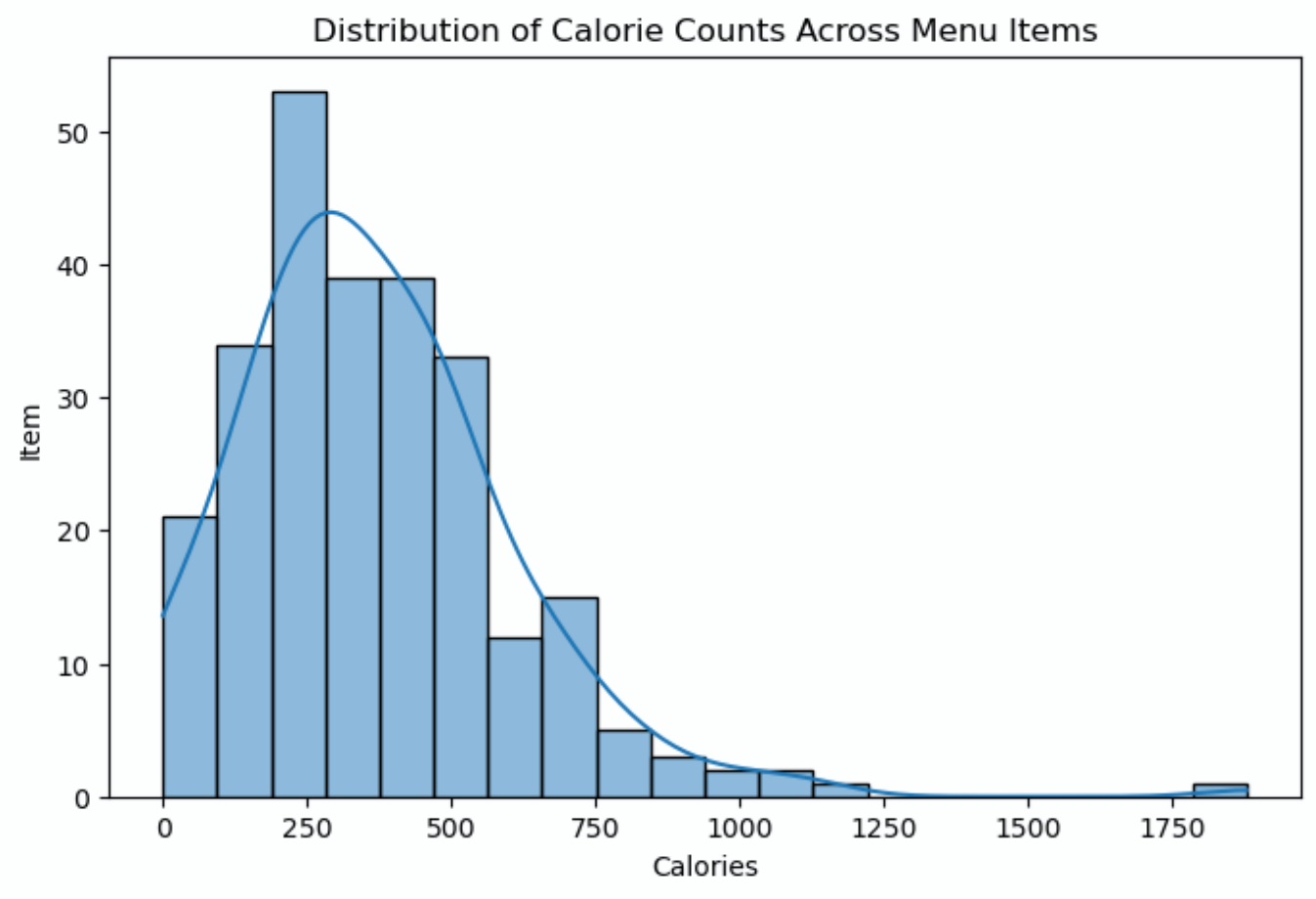
- ✓ **Figuring Out the Distribution:** Analyzing the distribution of nutritional variables within the dataset to understand the spread and central tendencies of the data.
- ✓ **Cleaning the Outliers:** Identifying and addressing outliers in the dataset to enhance the accuracy and reliability of subsequent analyses.
- ✓ **Plotting Correlation:** Creating correlation plots to visualize relationships between different nutritional components, providing insights into potential dependencies.
- ✓ **Data Visualization**: Employing data visualization techniques, such as histograms and box plots, to present a comprehensive overview of the nutritional content and identify patterns.
- ✓ **Nutrition-Based Insights:** Deriving meaningful insights from the data, including trends and patterns related to nutritional components, to inform decision-making.
- ✓ **Interactive Charts on Nutritional Info:** Utilizing interactive charting tools or libraries to create dynamic visualizations that allow users to explore nutritional information interactively, enhancing engagement and understanding.

Exploratory data analysis findings and Insights

- **Size of the Dataset :** The dataset has 260 entries (rows) and 24 columns.
- **Column Information :** There are three data types present in the DataFrame : float64, int64, and object (presumably strings or mixed types).
- **Numeric Columns Statistics :** Descriptive statistics for numeric columns (count, mean, std, min, 25%, 50%, 75%, max) are stating that there might be some severe outliers on all the nutritional columns
- **Categorical Columns :** The Category, Item, and Serving Size columns are of the object data type, indicating they are likely categorical or text data.
- **Nutritional Information :** Columns like Calories, Total Fat, Cholesterol, Sodium, Carbohydrates, Dietary Fiber, Sugars, and Protein provide nutritional information .The "% Daily Value columns" represent the percentage of daily recommended intake for various nutrients.
- **Data Completeness :** There are no missing values (Non-Null Count is 260 for all columns) and no duplicate values, indicating that the dataset is complete in terms of non-null entries.

Exploratory data analysis findings and Insights

Figuring out the distribution with a histogram :



A right-skewed distribution is also referred to as positively skewed

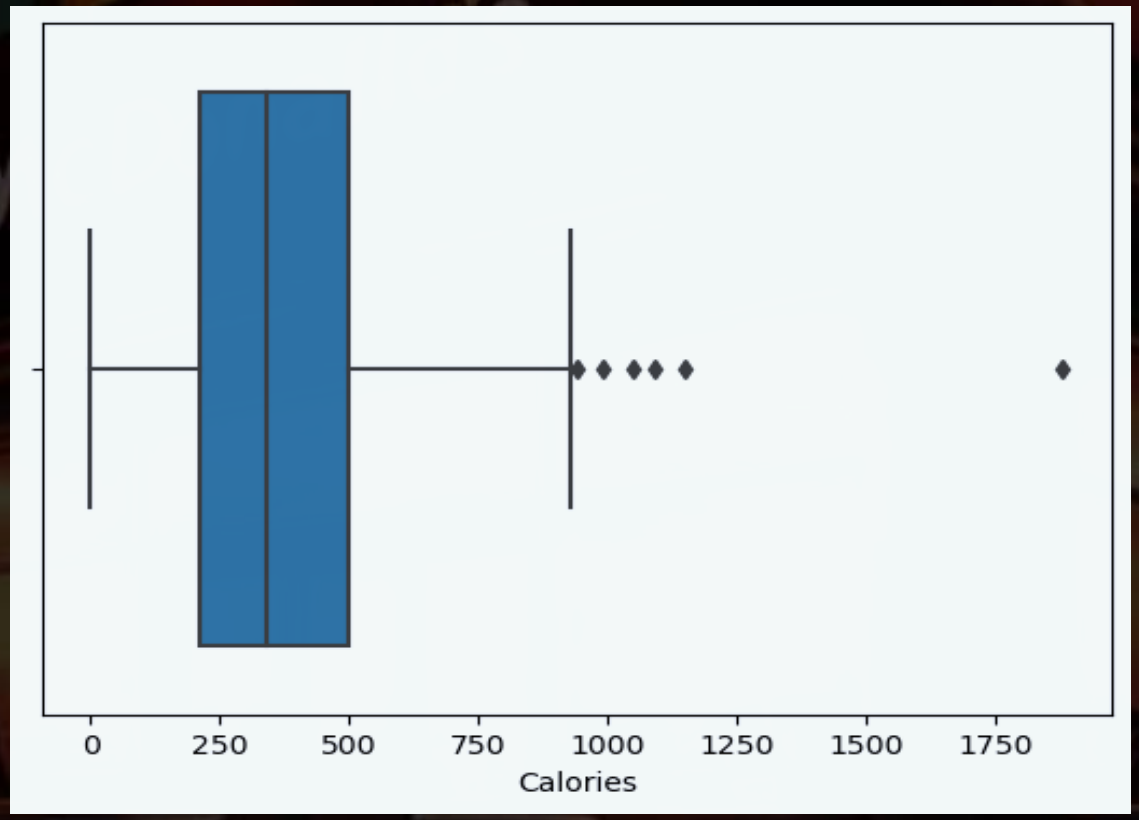
it implies that the majority of the data points are concentrated on the left side of the distribution, with a tail extending to the right The right tail of the distribution is longer, indicating the presence of relatively few data points with higher values.

These higher values contribute to the skewness and can affect the interpretation of central tendency.

So now we have to identify the outlier which causing the skewness and handle it

Exploratory data analysis findings and Insights

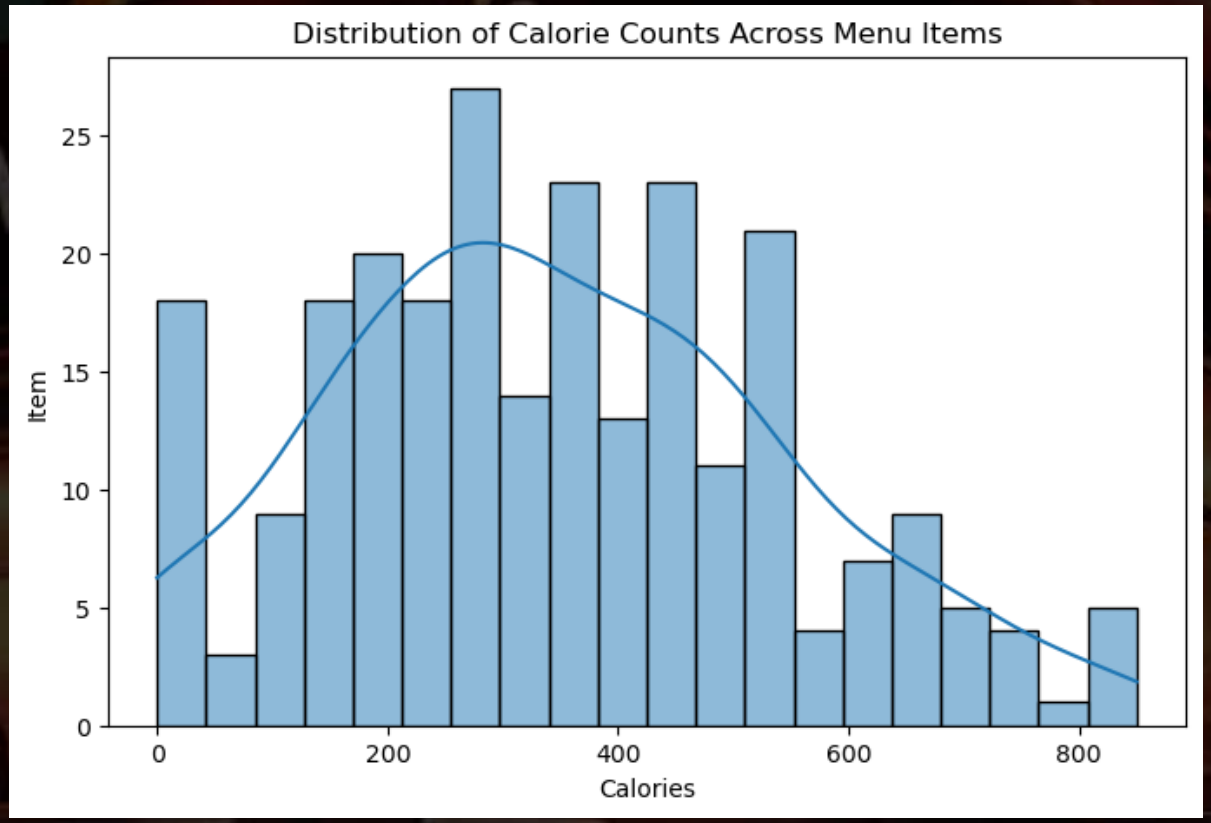
Finding outliers with a Box plot :



Upon inspecting the boxplot above, it is evident that there exist outliers in the dataset, specifically those exceeding the approximate threshold of 900 calories. Consequently, we will proceed to address and manage these outliers through appropriate techniques.

Exploratory data analysis findings and Insights

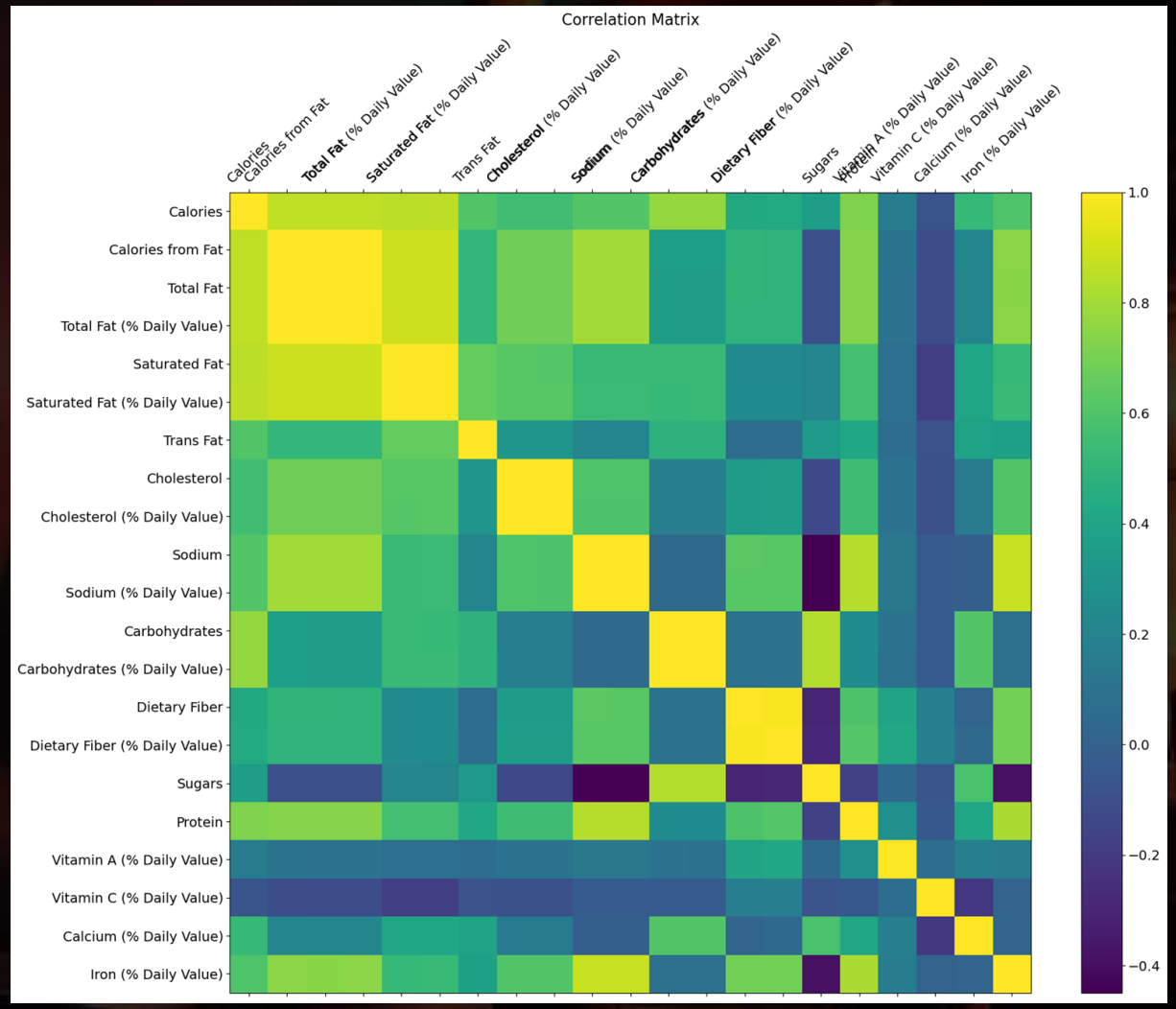
Distribution after cleaning the outliers



From the above Figure, we can Observe a bell-shaped curve after removing outliers indicates a positive outcome in terms of normalizing the distribution. However, it's essential to carefully consider the impact of outlier removal on our analysis

Exploratory data analysis findings and Insights

Correlation Matrix:



The correlation coefficients are proportionate to the box's size and color intensity.

As we can see, the variable (Calories), has strong, obvious correlations with all of the other independent variables, whether they are positive or negative.

Data Visualizations :

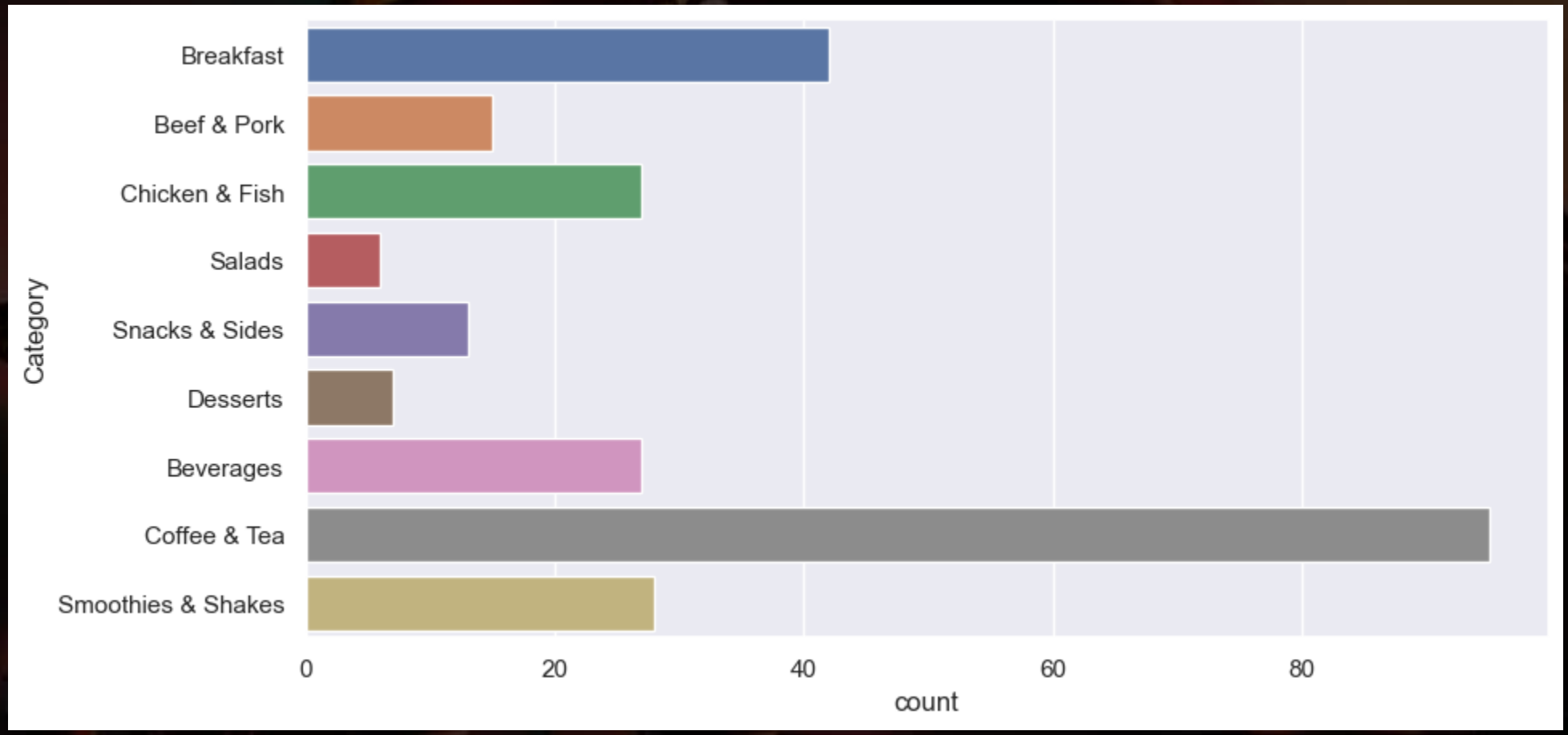
In this data visualization section, diverse aspects of the data are meticulously analyzed and presented visually to ensure a lucid comprehension.

This approach allows for a comprehensive understanding of the nutritional landscape, facilitating insights that contribute to informed decision-making and potential improvements in the realm of fast-food nutrition

- No .of dishes available in different categories
- Displaying distribution of **Calories** over categories
- Displaying distribution of **Cholesterol** over categories
- Displaying distribution of **Fat** and **Saturated fat** over categories
- Displaying distribution of **Vitamin A** and Vitamin C over categories
- Displaying distribution of **Calcium** over categories
- Displaying distribution of **Iron** over categories
- Displaying distribution of **Carbohydrates** over categories
- Displaying distribution of **Dietery fibers** over categories

Data Visualizations :

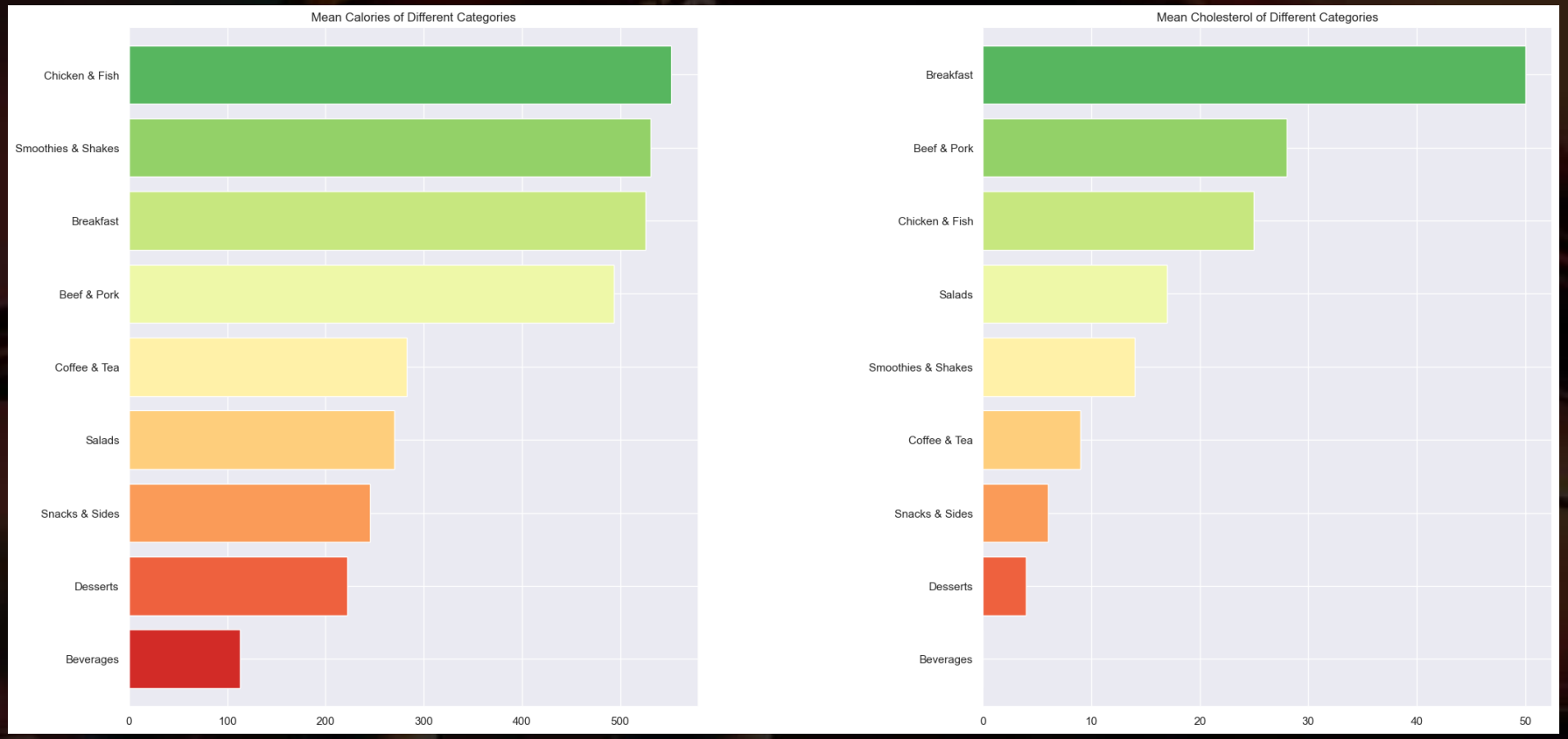
Displaying number of menus in specific categories :



From the above , we can understand that Coffee & Tea category have more items than other categories

Data Visualizations :

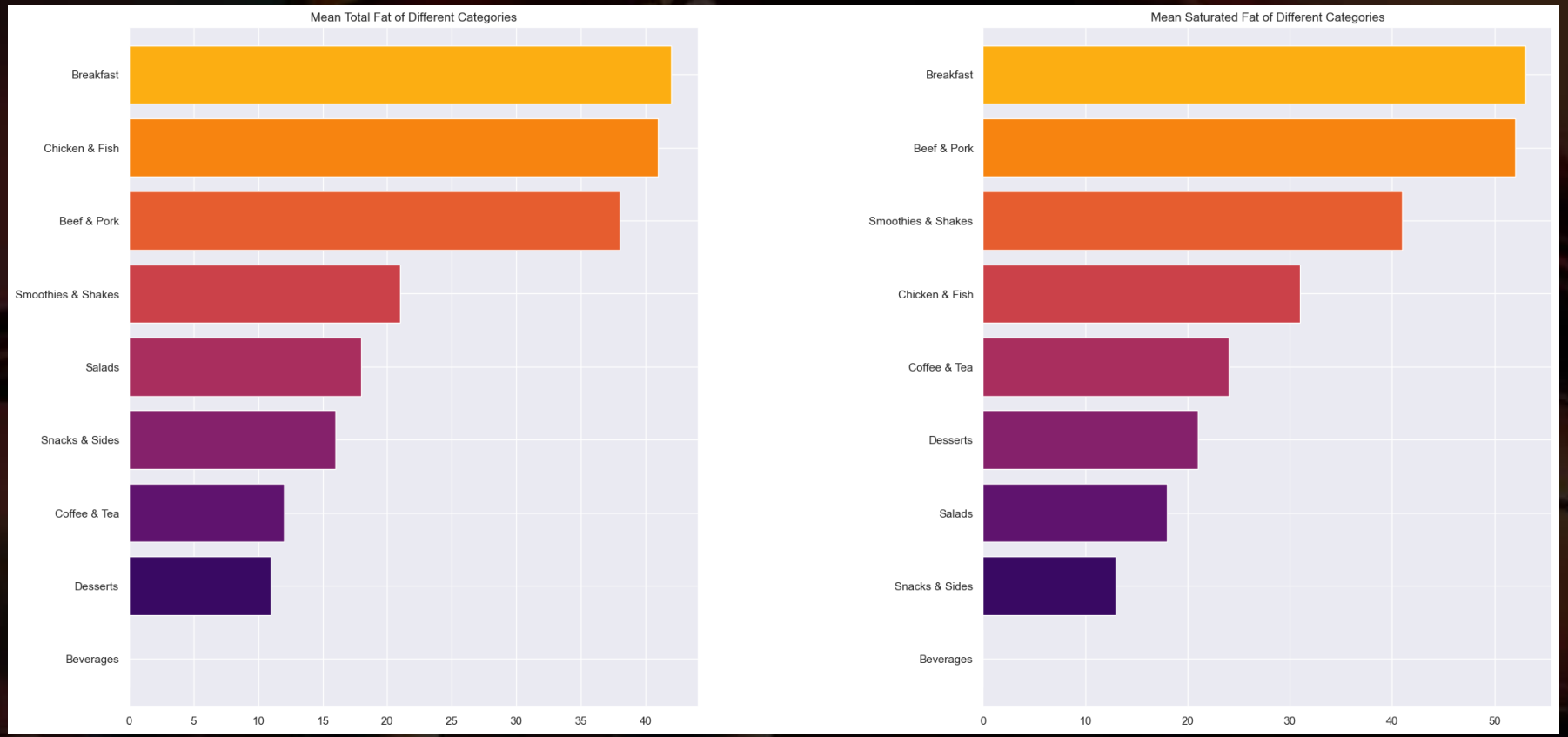
Displaying Distribution of Calories & Cholesterol in specific categories :



From the above diagram we can understand the high calorie and high cholesterol food categories.

Data Visualizations :

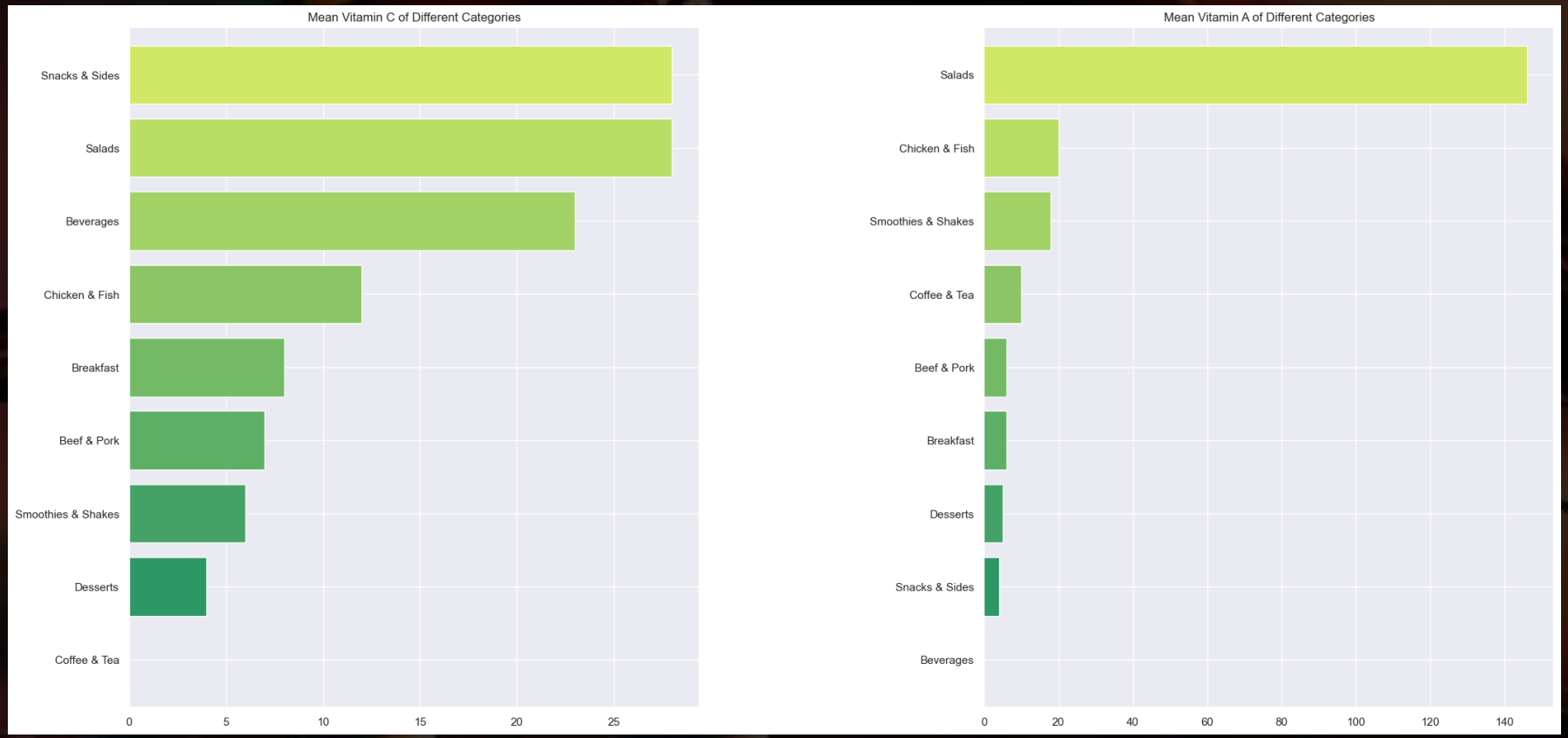
Displaying Distribution of Total Fat & Saturated Fat in specific categories :



From the above diagram we can understand the high Total Fat and Saturated Fat food categories

Data Visualizations :

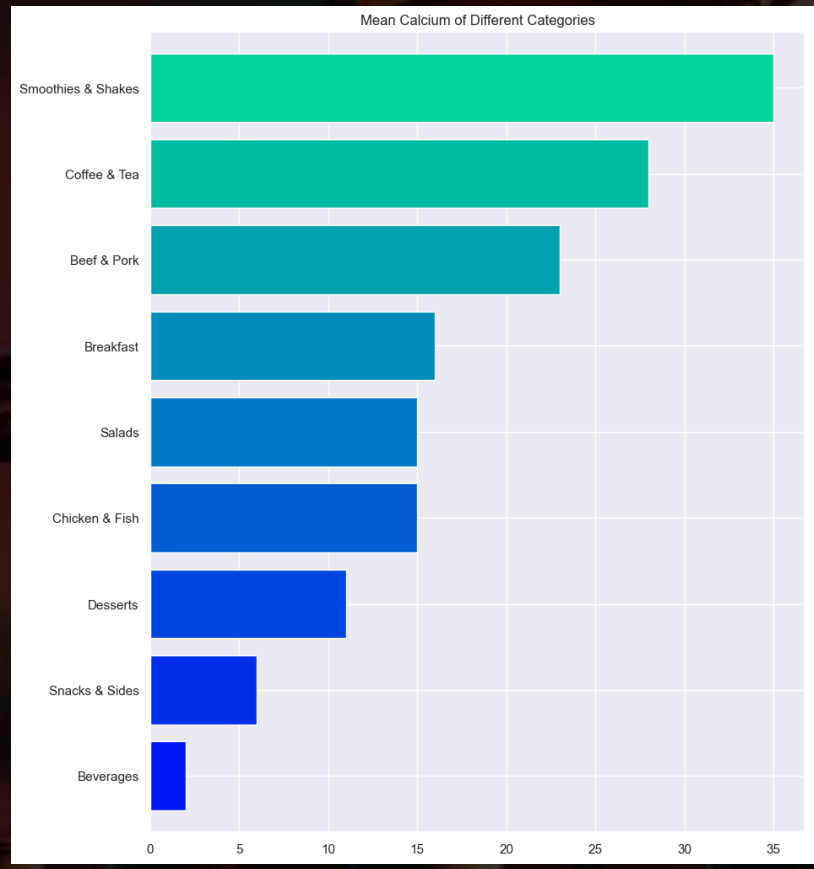
Displaying Distribution of Vitamin C & Vitamin A in specific categories :

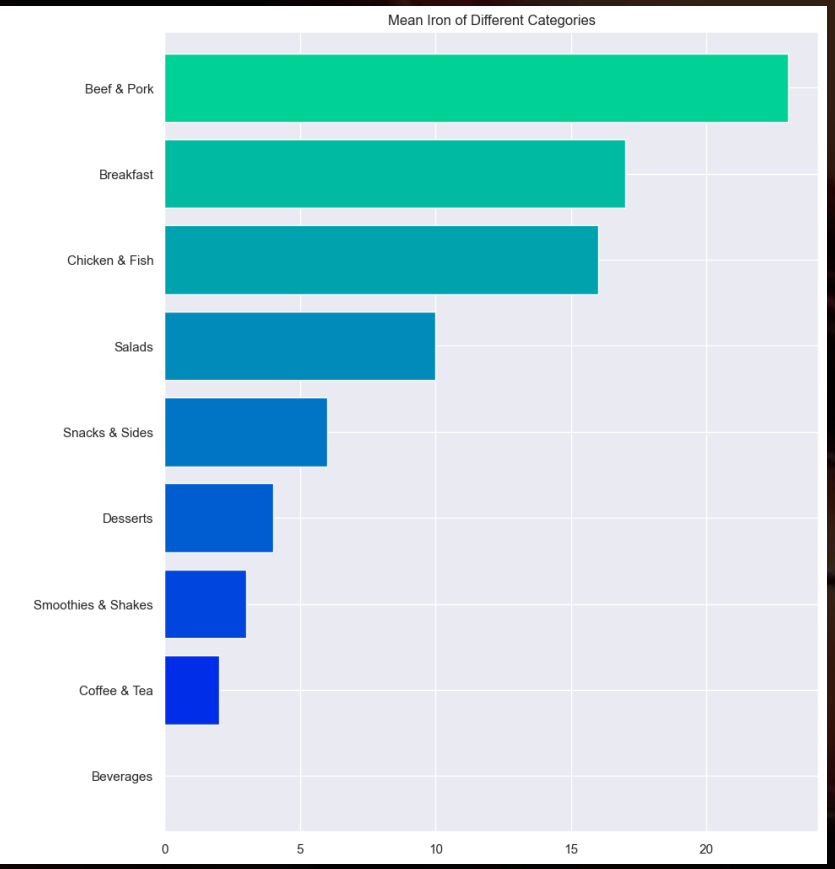


From the above diagram we can understand the Vitamin C and Vitamin A Rich food categories

Data Visualizations :

Displaying Distribution of Calcium & Iron in specific categories :

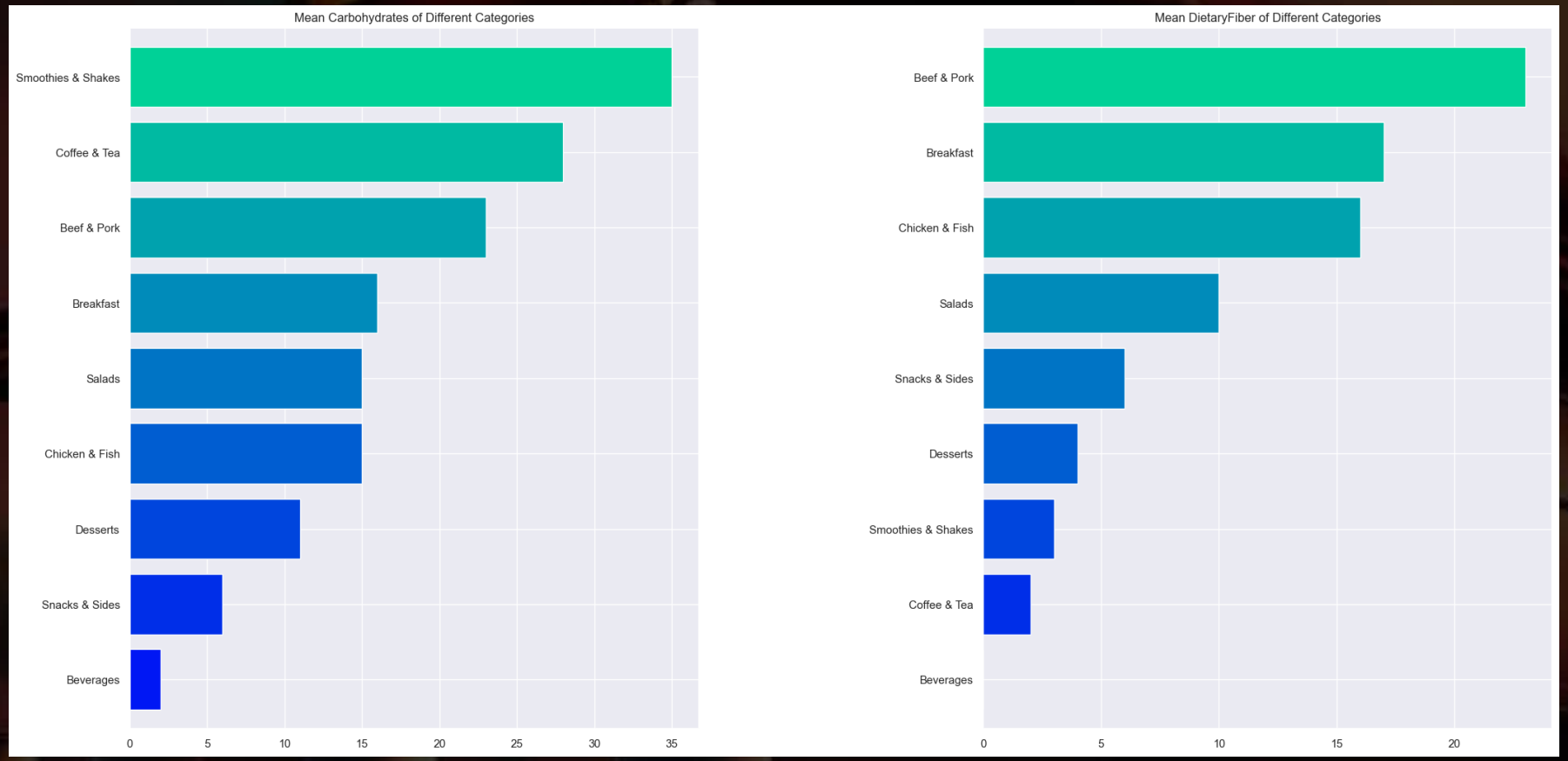




From the above diagram we can understand the Calcium and Iron Rich food categories

Data Visualizations :

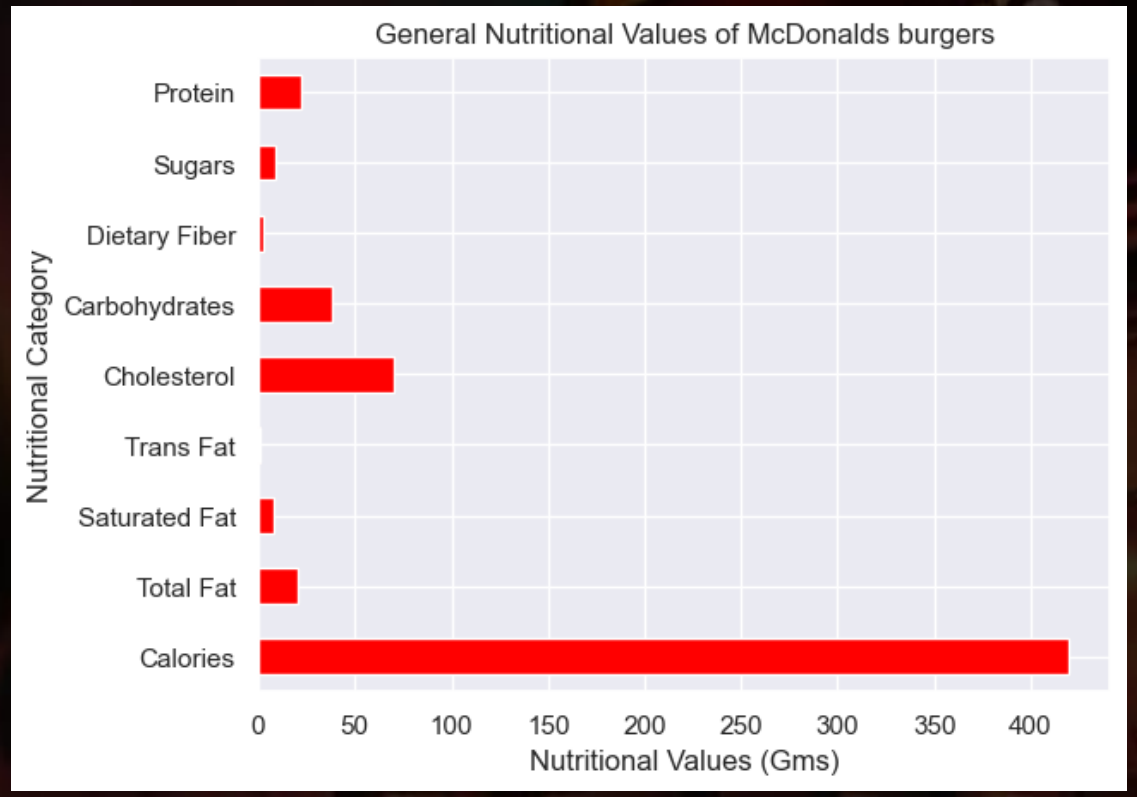
Displaying Distribution of Carbohydrates & DietaryFiber in specific categories :

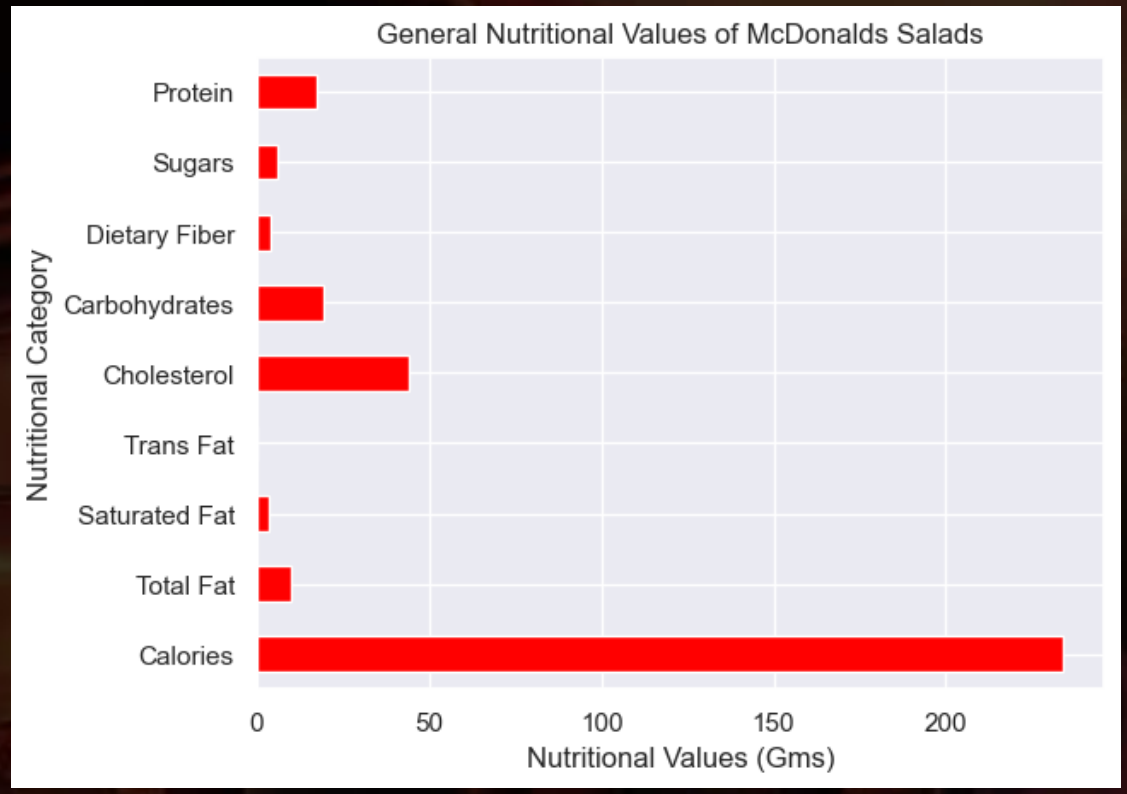


From the above diagram we can understand the high carb food category and also food categories which are rich in food category

Data Visualizations :

Comparing Nutritional facts between all burgers and salads in McDonalds :

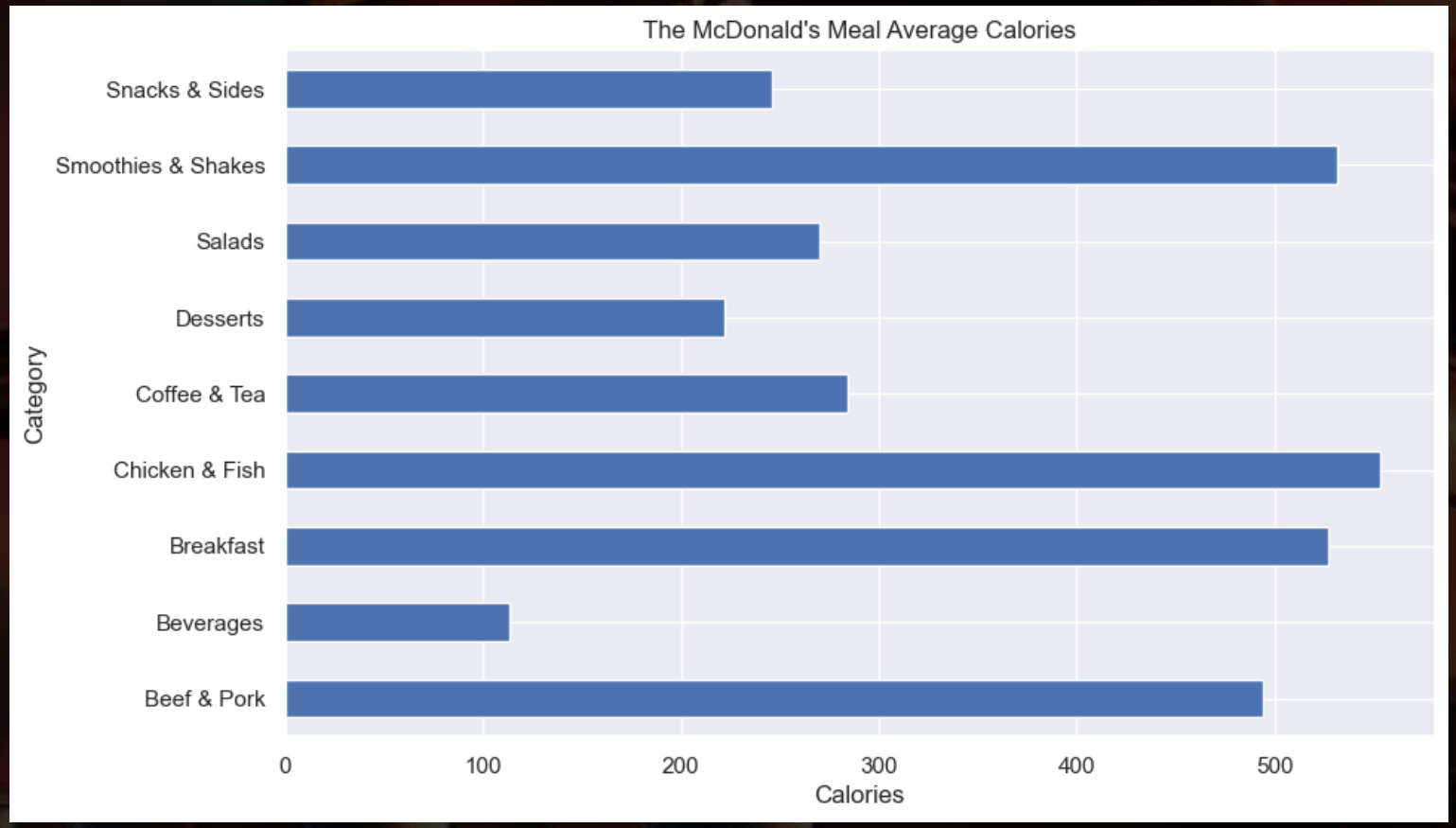




From the above diagram we can clearly understand the nutritional facts difference between burgers and salads of McDonalds.

Nutritional Based Insights :

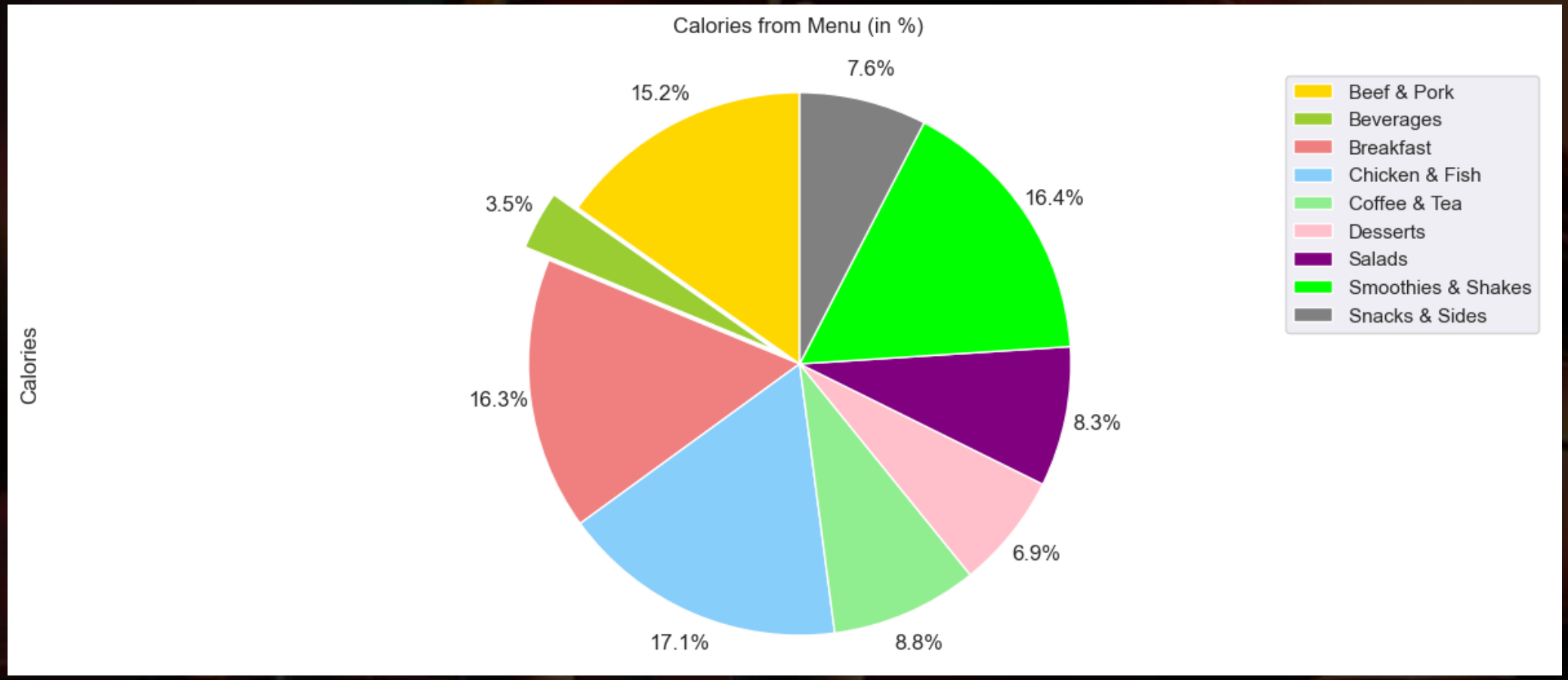
McDonald’s meals average calories over categories



From the above diagram we can clearly understand the calories distribution over food categories

Nutritional Based Insights :

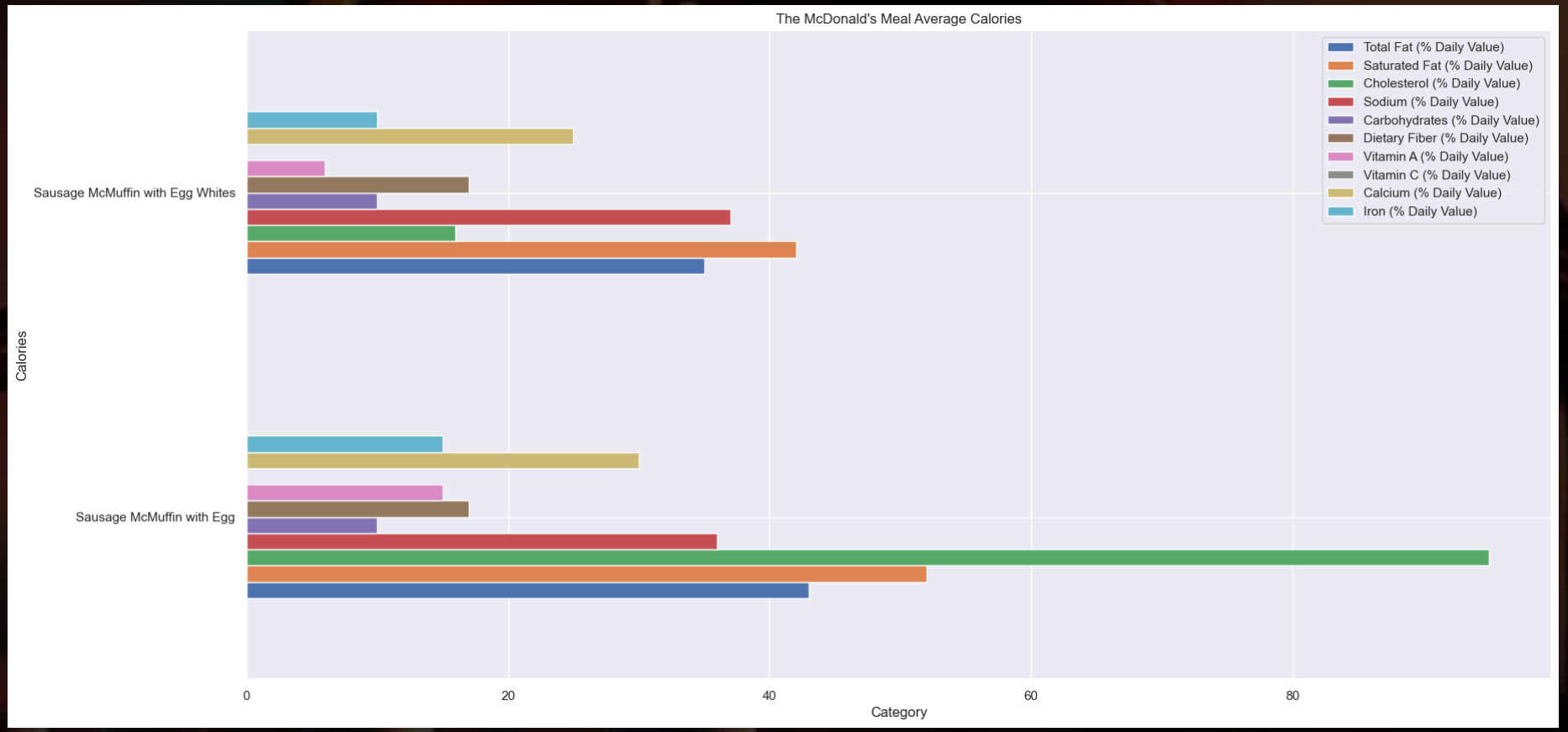
Overall caloric intake of all categories and beverage's contribution in that



From the above Pie chart ,we can see that beverage only contribute 3.2% to overall caloric intake.

Nutritional Based Insights :

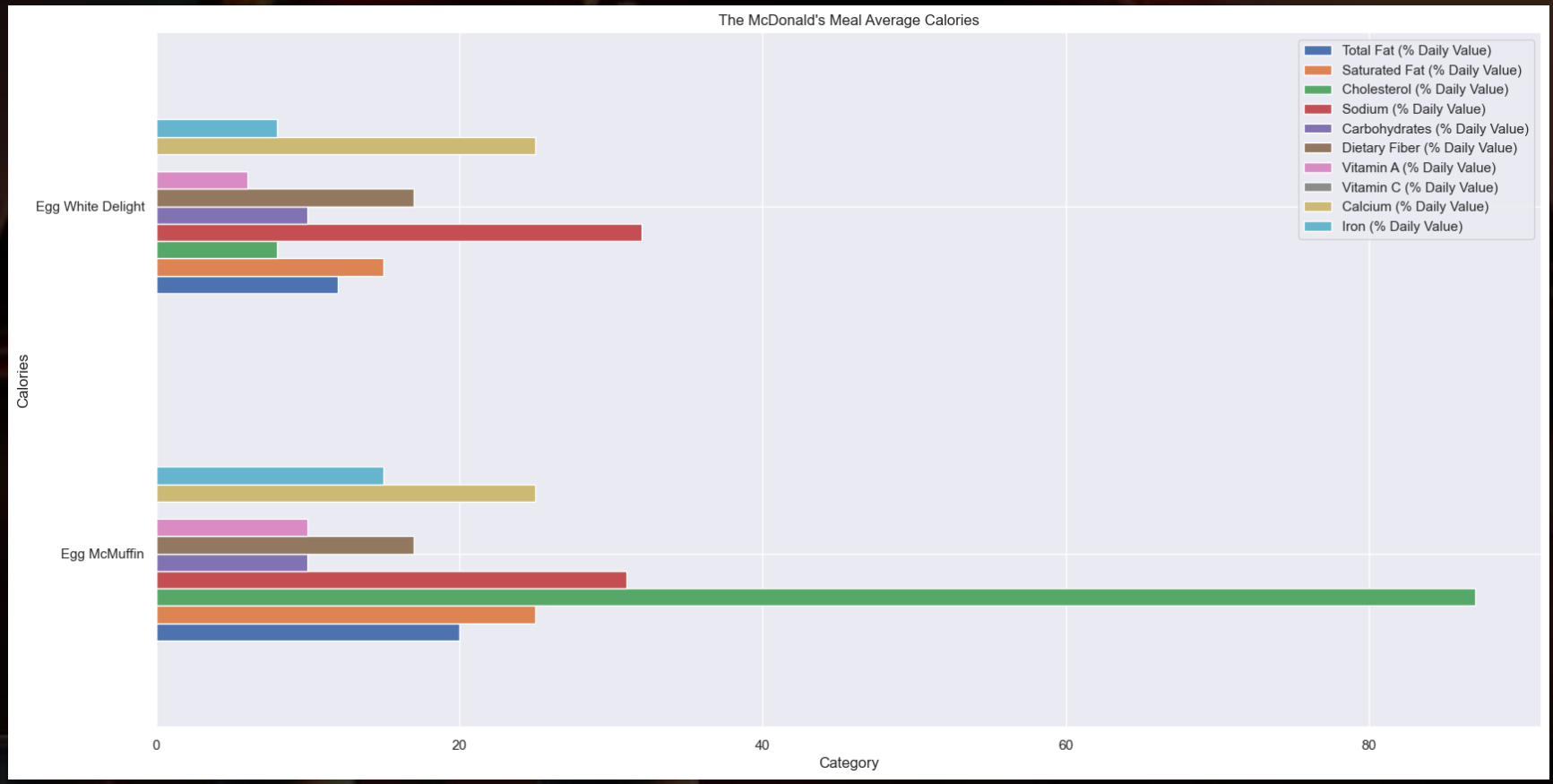
(1) Comparing two dishes that contain whole egg and only egg white



From the above observation we can understand there is significant difference in terms of nutrition between Only egg white and whole egg

Nutritional Based Insights :

(2) Comparing another two dishes that contain whole egg and only egg white



From the above observation we can understand there is significant difference in terms of nutrition between Only egg white and whole egg

Nutritional Based Insights :

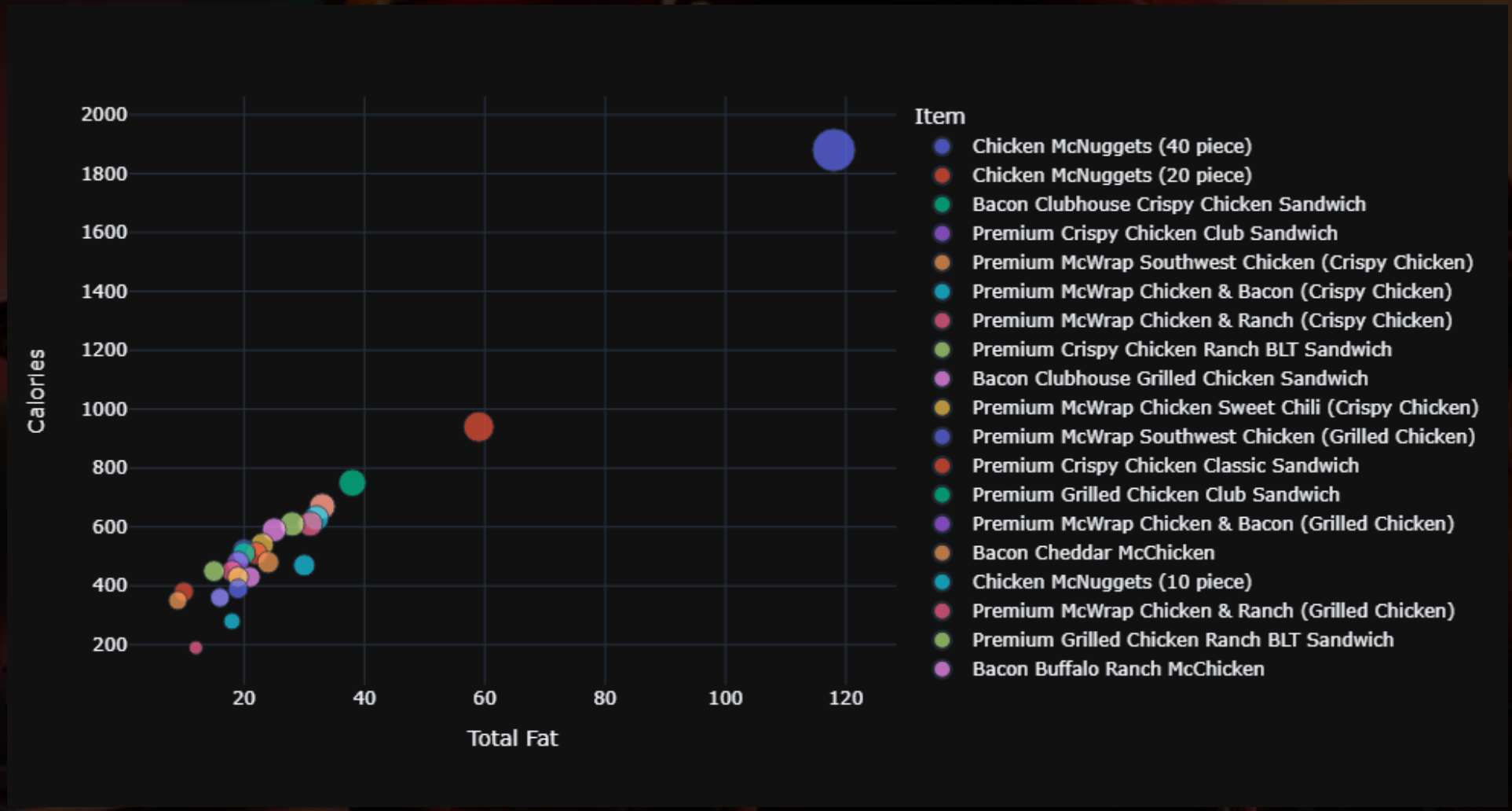
These are the 10 high and low calorie (non zero) foods of Mcdonalds:

Item	Calories
Chicken McNuggets (40 piece)	1880
Big Breakfast with Hotcakes (Large Biscuit)	1150
Big Breakfast with Hotcakes (Regular Biscuit)	1090
Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)	1050
Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit)	990
Chicken McNuggets (20 piece)	940
McFlurry with M&Mâ€™s Candies (Medium)	930
Strawberry Shake (Large)	850
Chocolate Shake (Large)	850
Vanilla Shake (Large)	820

Item	Calories
Apple Slices	15
Side Salad	20
Kids Ice Cream Cone	45
Minute Maid 100% Apple Juice Box	80
Iced Coffee with Sugar Free French Vanilla Syrup (Small)	80
Coca-Cola Classic (Child)	100
Dr Pepper (Child)	100
Sprite (Child)	100
1% Low Fat Milk Jug	100
Nonfat Latte (Small)	100

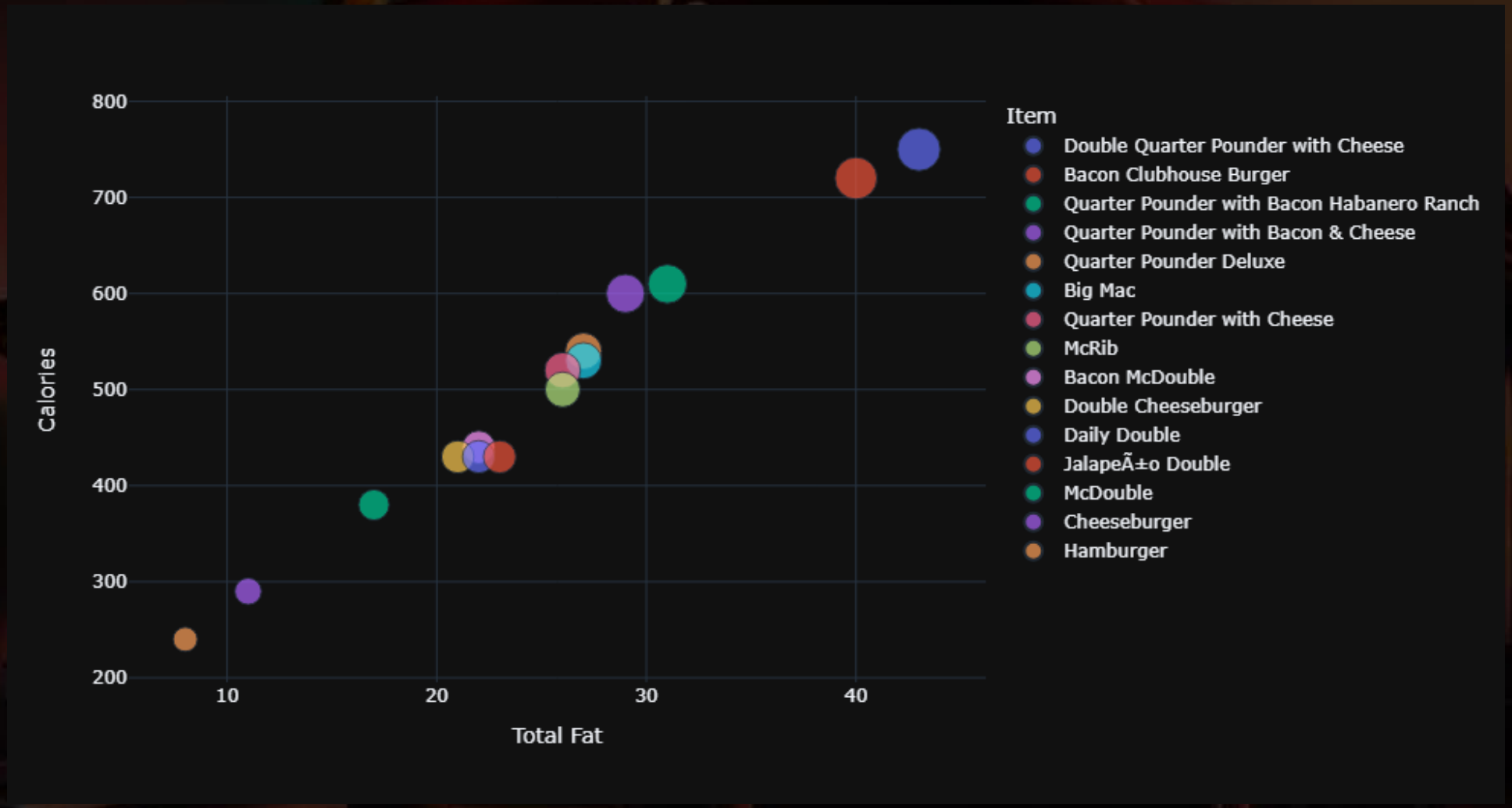
Nutritional Based Insights :

Interactive plot displaying Total fat and Calories of “Chicken and Fish” category



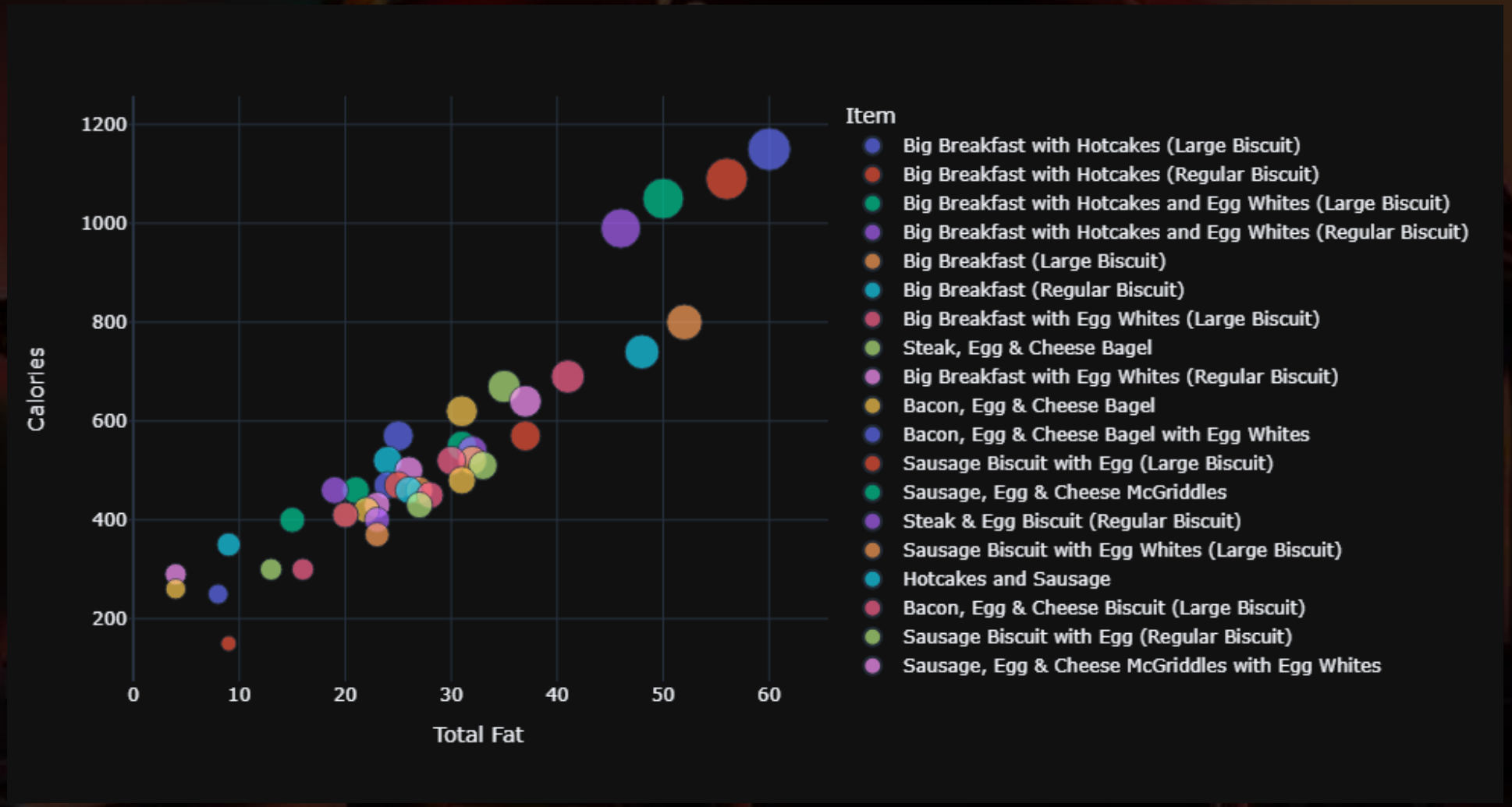
Nutritional Based Insights :

Interactive plot displaying Total fat and Calories of “Beef and Pork” category



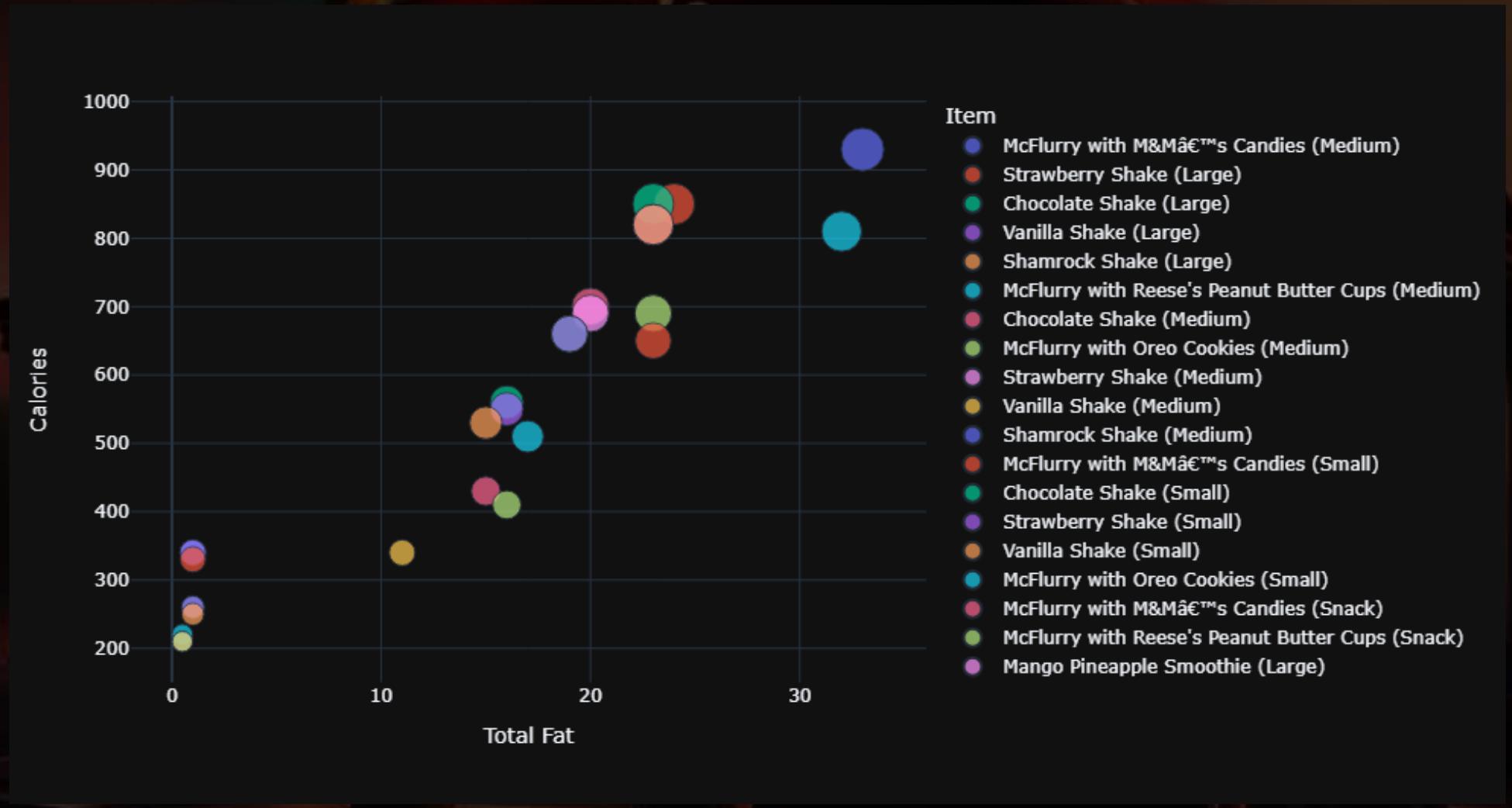
Nutritional Based Insights :

Interactive plot displaying Total fat and Calories of “Breakfast” category



Nutritional Based Insights :

Interactive plot displaying Total fat and Calories of “Smoothies & Shakes” category



Summarization of analysis:

Menu Distribution : The "Coffee & Tea" category dominates with the highest number of food items, followed by "Chicken and Fish" and "Smoothies and Shakes.”

Caloric Insights :"Chicken and Fish" items have the highest calorie content, followed by "Smoothies and Shakes" and "Breakfast," while "Burgers" tend to be calorically dense.

Cholesterol Concerns :"Breakfast" items exhibit elevated cholesterol levels compared to other categories.

Vitamin-Rich Salads :"Salads" stand out as vitamin-rich, particularly in Vitamin A and Vitamin C.

Nutrient-Dense Beef and Pork :"Beef and Pork" dishes excel in iron content and dietary fibers.

Summarization of analysis:

Calcium Conundrum in Shakes :"Smoothies and Shakes" boast high calcium but also high carbs, showcasing a nutritional trade-off.

Burgers vs. Salads : Comparatively, burgers are calorically high but nutrient-poor, while salads offer low calories with high nutritional value.
Beverages Contribution:

Surprisingly, "Beverages" contribute only 3.2% to the overall calorie content of McDonald’s.

Egg Options Analysis : Whole eggs showcase rich nutrients but higher cholesterol, whereas egg whites offer similar nutrients with lower cholesterol.

Conclusion :

- **According to FDA**, the recommended adult limit for sodium intake should be less than **2,300** milligrams per day - that's equal to about **1 teaspoon** of table salt.
- **FDA** also recommends adults consume **50 grams (g)** of protein a day.
- **USDA** recommends 66-97 grams of fat per day in a **2,000** and **2,500** calorie diet respectively.
- **NHS** recommends daily calorie intake of **2,000 kcal for women and 2,500 kcal for men**.
- **'Chicken McNuggets (40 piece)'** was found to contain high amounts of Calories **(1880 kcal), Sodium (3,600 milligrams), Protein (87 grams),** and **Total Fat (118 grams).**

In summary, this analysis unveils intricate details about McDonald's menu items, emphasizing the need for informed choices based on individual nutritional preferences. Whether it's maximizing nutrients or managing caloric intake, understanding these nuances empowers consumers to make healthier decisions aligned with their dietary goals.