

Analyzing Nutrition and Calorie Intake from Fast Foods

2024-12-1

L-01 Group 9:

Itunuoluwa Olowoye (UCID: 30184712)
Eisha Imran Chaudhry (UCID: 30263099)
Vrinda (UCID: 30270428)
Israel Maikyau (UCID: 30269917)

DATA 603 Project Report

1. Introduction

1.1. Motivation

1.1.1. Context

This project focuses on examining the nutritional value of fast food using menu items from selected fast food restaurants as samples. Eating food away from home (FAFH), particularly fast food, is a prevalent practice in Canada and is associated with poorer dietary habits. This includes lower intakes of essential nutrients such as fruits, vegetables, and whole grains, while there is higher consumption of fats, sodium, and sugars (Polksy & Garriguet, 2021). This makes understanding the nutritional content of fast food items more crucial, especially considering the broader societal challenges related to diet and health highlighted by Marcone et al. (2020).

1.1.2. Problem

How can we make informed dietary choices when it comes to fast food? Fast food restaurants offer easily accessible meals, but they also come with concerns about their high calorie and processed fat content. As evident, the nutritional quality of home-cooked meals is not reflected in that of fast food, and the choice of eating fast food is influenced by socio-environmental factors like affordability, accessibility, and convenience (Marcone et al., 2020). Considering some of these factors are unavoidable, addressing these issues with informed decisions is a public health priority, given the increasing reliance on fast food.

1.1.3. Challenges

Even though fast food is a convenient option, its frequent consumption can lead to several health issues, such as obesity, diabetes, and heart disease. In Canada, approximately 1 in 8 households experience food insecurity, which can lead to lower-cost food options that contribute to poor nutritional outcomes (Marcone et al., 2020). Public health discussions about the long-term impacts of fast food are necessary to guide policy changes aimed at improving nutritional standards within the fast-food industry.

1.2. Objectives

1.2.1. Overview

While there has been a growing recognition of the health risks that are associated with fast food consumption, there still remains a significant gap in addressing food insecurity and making nutritious food more accessible and affordable (Marcone et al., 2020). By understanding the relationship between nutrients and calorie content, the project aims to give consumers more transparency and help them make healthier choices when selecting fast food menu items.

1.2.2. Goals & Research Questions

The goal is to construct a robust predictive model that can help quantify the calorie value of fast food based on commonly known nutrients. This project aims to provide useful insights which could inform initiatives aimed at improving transparency within the fast-food industry, leading to the development of healthier menu options.

For this project, we will be exploring three research questions:

1. What is the calorie content distribution for different food categories and in different fast food restaurants?
2. Trans Fats form when hydrogen is added to vegetable oil. This is done to increase shelf life but is bad for our health and the FDA is forcing food companies to phase them out. Is there a significant difference in the average trans fat content of menu items in the selected fast food restaurants?
3. Can we establish a linear relationship between calories and the nutritional content of various items on the menu of different fast food restaurants?

2. Methodology

2.1. Data

The dataset used in this project, the Fast Food Nutrition Dataset, was web-scraped from the Fast Food Nutrition Facts website (Fast Food Nutrition Facts, n.d.). This dataset was originally available on Kaggle (Arvidsson, 2023), but some of the rows in the data contained misrepresentations of values (also noticed by other individuals in the discussion), which we considered undesirable for our analysis.

The data was collected through a web scraping method and is representative of a sample of fast food restaurants in Canada. This means the sample is not randomly selected but rather manually chosen. This also introduces a potential selection bias as it limits the analysis to specific restaurants that may not be fully representative of the entire fast food industry in Canada.

Each row in the dataset provides a detailed breakdown of the nutrient and calorie values of menu items from the selected fast-food restaurants. The geographical scope of the data is limited to the Canada. It includes **745 rows and 28 columns**, out of which we will use **24 columns** comprising **23 predictor variables** and **1 response variable**. The predictor variables are described in **Table 1** below.

Table 1

Predictor Variables

S/N	Predictor Variable	Variable Type	Unit
1	Company	Qualitative	No units
2	Calories From Fat	Quantitative	kcal
3	Total Fat (g)	Quantitative	grams
4	Saturated Fat (g)	Quantitative	grams
5	Trans Fat (g)	Quantitative	grams
6	Cholesterol (mg)	Quantitative	milligrams
7	Sodium (mg)	Quantitative	milligrams
8	Total Carbohydrates (g)	Quantitative	grams
9	Dietary Fiber (g)	Quantitative	grams
10	Sugars (g)	Quantitative	grams
11	Protein (g)	Quantitative	grams
12	Total Fat (% Daily Value)	Quantitative	percentage
13	Saturated Fat (% Daily Value)	Quantitative	percentage
14	Cholesterol (% Daily Value)	Quantitative	percentage
15	Sodium (% Daily Value)	Quantitative	percentage
16	Total Carbohydrates (% Daily Value)	Quantitative	percentage
17	Dietary Fiber (% Daily Value)	Quantitative	percentage
18	Protein (% Daily Value)	Quantitative	percentage
19	Vitamin A (% Daily Value)	Quantitative	percentage
20	Vitamin C (% Daily Value)	Quantitative	percentage
21	Calcium (% Daily Value)	Quantitative	percentage
22	Iron (% Daily Value)	Quantitative	percentage
23	Category	Qualitative	No units

Note. This table presents the predictor variables, their variable types, and units that will be used in the analysis.

The *Response Variable* in the scope of this project is **Calories** (Quantitative, calories).

2.2. Approach

For the research questions about calorie content distribution for different food categories and in different fast food restaurants, we will visualize with bar charts and box plots the difference between the calorie content of different companies and different food categories. We will do the same with understanding the difference in the average trans fat content of menu items in the selected restaurants.

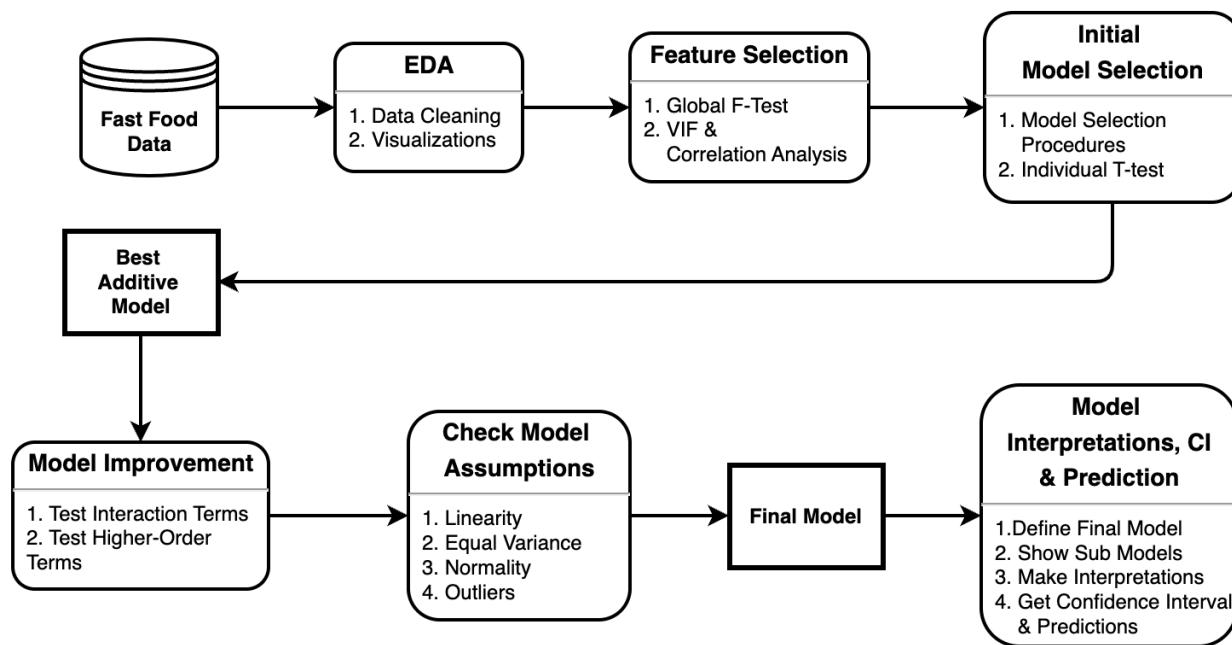
For establishing a linear relationship between calories and nutritional content, we will build a **multiple linear regression model** between calories and the different nutrients.

2.3. Workflow

1. We load the required R libraries, load and clean the data
2. We visualize the average calories of different food categories
3. We visualize the calorie distribution of different fast food restaurants
4. We examine the trans fat content of menu items from these different fast food restaurants
5. We build a *multiple linear regression model* of the effect of nutrients, company, and food category on calories, using the following steps. We are employing this method because the variable of interest, calories, is quantitative and we want to assess whether there are relationships between this variable and many other variables (nutrients, company, food category):
 - Feature Selection: Define dependent variables, predictor variables, and assess multicollinearity between predictor variables using the variance inflation factor (VIF).
 - Perform Stepwise Selection, Forward, Backward and All-Possible-Regressions Selection Procedures to select a starting model.
 - Verify the starting model using the Individual T-test. The Individual T-test helps to calculate individual p values for predictors we can compare with chosen significance level. Our **significance level $\alpha = 0.05$**
 - Perform model improvement tasks such as including interaction and higher-order terms to improve the model.
 - Select a valid model for prediction with proof of why this is the best choice.
 - Validate model assumptions (Linearity, equal variance and normality assumptions) and detect (or remove) any outliers that may significantly affect the model.
 - Perform model transformations, if needed, when the model assumptions are violated.

Figure 1

Workflow Diagram



Note. This workflow diagram outlines the key steps involved in the model development process for the fast food nutrition analysis project.

2.4 Contributions

1. Eisha:

EDA (Exploratory Data Analysis)

- Data cleaning: Handle missing values, detect outliers, and check data types.
- Generate summary statistics: Summary of all variables.
- Visualize distributions: Histograms, box plots, scatter plots, and pair plots.
- Check correlations: Between numeric variables.

2. Israel:

Feature Selection

- Correlation analysis: Identify highly correlated features.
- Feature importance ranking: Use domain knowledge and EDA
- Statistical methods: Feature selection based on p-values.

3. Vrinda:

Full Additive Model and Partial Model Tests

- Develop a full regression model: Use all selected features.
- Test partial model: Use a reduced set of features.
- Generate predictions: Compare model performance.

4. Itunuoluwa:

Interaction Model or Higher-Order Model, Model Selection & Validation

- Interaction model or Higher order model: Build a model with interaction or higher-order terms.
- Select best model for prediction from valid models.
- Validate model assumptions.
- Interpret selected model, if interpretation is not lost.

3. Main Results of the Analysis

First the required R libraries and the dataset are loaded. Then we proceed to clean the data.

Table 2

Head of Dataset

Company	Item	Calories	Calories.From.Fat	Total.Fat..g	Saturated.Fat..g
a-w	Baby Burger without Cheese	240	72	8	3.0
a-w	Baby Burger with Cheese	298	117	13	6.0
a-w	Beyond Meat Burger	500	261	29	5.0
a-w	Buddy Burger without Cheese	310	117	13	4.5
a-w	Buddy Burger with Cheese	370	144	16	7.0
a-w	Cheddar Bacon Uncle Burger	760	432	48	19.0

Note. This table presents the first six rows of a dataset that includes a few columns, showcasing how the variable names are currently stored in the data.

The column names are converted to snake case and then we validate the calories data by filtering out any item whose calorie value is greater than the calories from fat. The new column names are shown in the R output below.

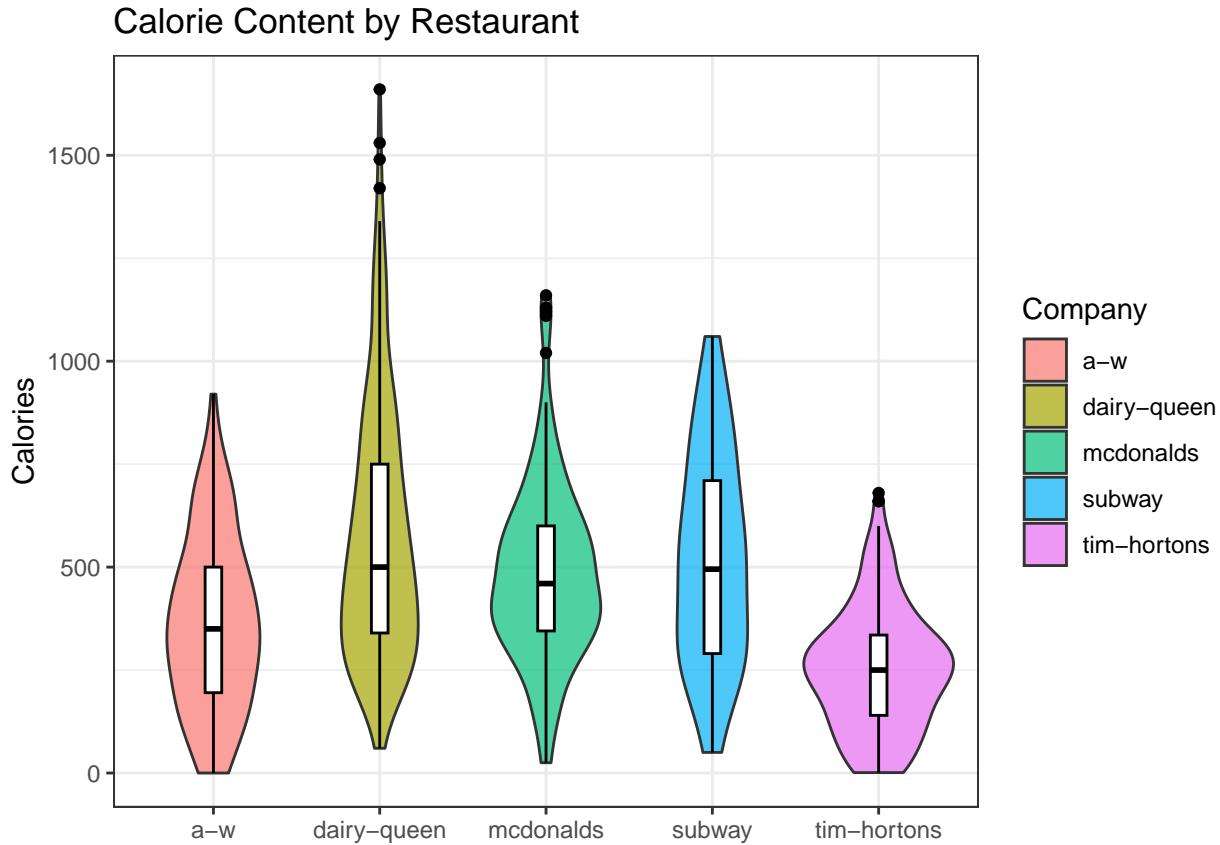
```
## The columns in the data include:
##  [1] "company"
##  [3] "calories"
##  [5] "total_fat_g"
##  [7] "trans_fat_g"
##  [9] "sodium_mg"
## [11] "dietary_fiber_g"
## [13] "protein_g"
## [15] "saturated_fat_daily_value"
## [17] "sodium_daily_value"
## [19] "dietary_fiber_daily_value"
## [21] "vitamin_a_daily_value"
## [23] "calcium_daily_value"
## [25] "category"
## [27] "smart_points"
##  [1] "item"
##  [2] "calories_from_fat"
##  [3] "saturated_fat_g"
##  [4] "cholesterol_mg"
##  [5] "total_carbohydrates_g"
##  [6] "sugars_g"
##  [7] "total_fat_daily_value"
##  [8] "cholesterol_daily_value"
##  [9] "total_carbohydrates_daily_value"
## [10] "protein_daily_value"
## [11] "vitamin_c_daily_value"
## [12] "iron_daily_value"
## [13] "free_style_points"
## [14] "points_plus"
```

3.0.1. Calorie content distribution in different fast food restaurants

First we aim to assess Calorie content distribution in different fast food restaurants. we do this by plotting a box plot and violin plot to see the distributions based on each company (using the ggplot library).

Figure 2

Violin Plot & Box plot of Calorie Content Distribution



Note. This figure presents a violin plot and box plot of the calorie content distribution for different fast food restaurants. The x-axis shows the different restaurant companies, and the y-axis displays the calorie content in calories.

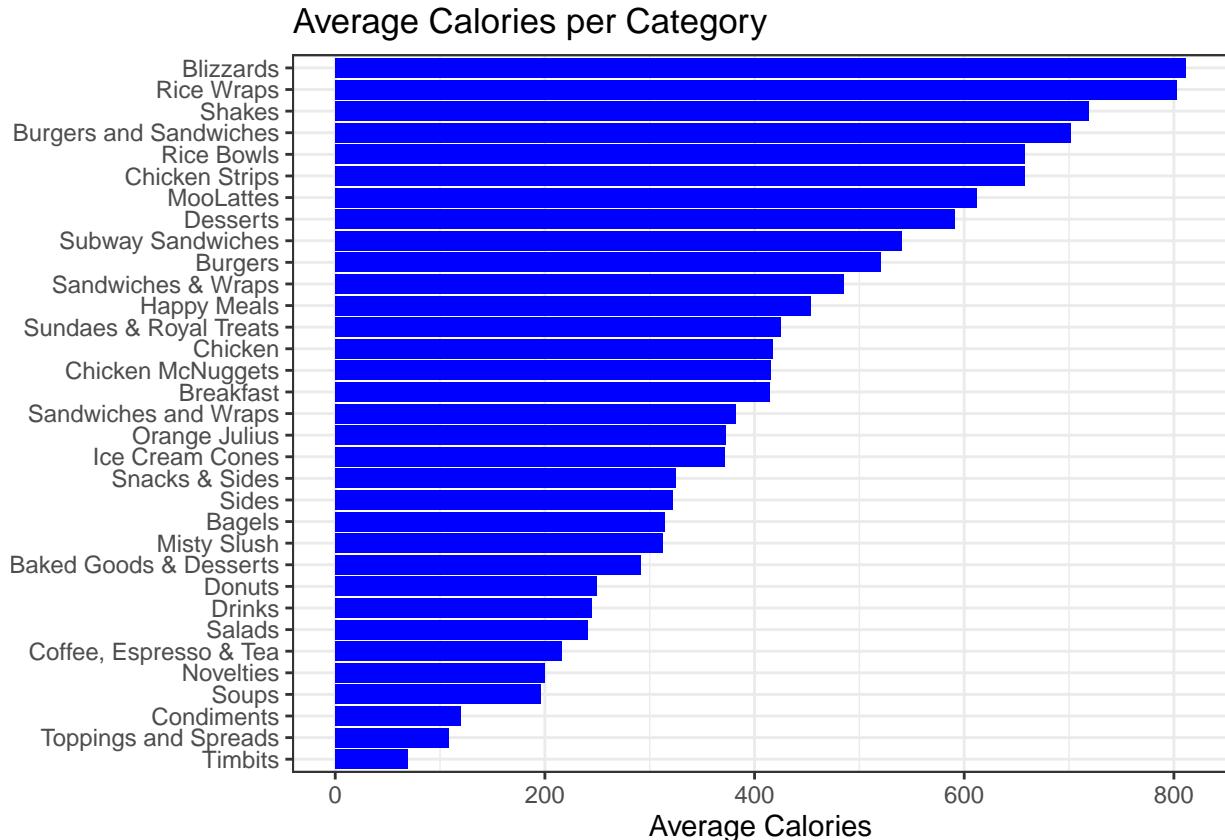
Diary Queen has certain menu items with more calories compared to the other fast food restaurants, causing its calorie distribution to be skewed to the right compared to others. In terms of median calorie content, Tim Hortons has the lowest, followed by A&W, McDonalds, Subway, and Dairy Queen respectively.

3.0.2. Average Calories Content of different food categories

Next we aim to get the Average Calories Content of different food categories. We calculate the average calorie content for each food category using the dplyr library, sort the categories by their average calorie values, and visualize the results using a horizontal bar chart created with ggplot to highlight the differences in calorie content across categories.

Figure 3

Average Calories Content per Category



Note. This figure presents the average calorie content per menu category for a restaurant or food provider. The categories are displayed vertically, and the corresponding average calorie content is shown on the horizontal axis.

Blizzards and Rice Wraps contain more than 800 calories, on average. Timbits contain the lowest calories on average.

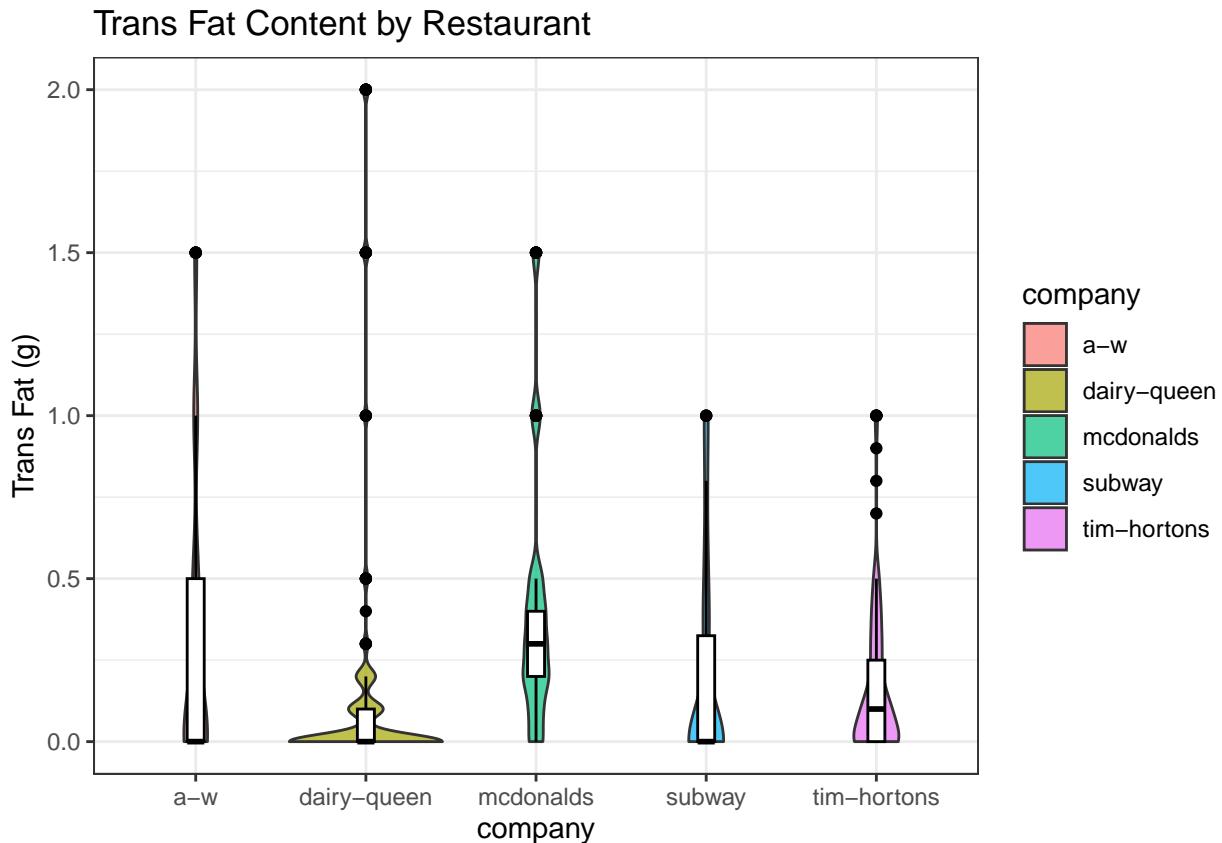
There are some categories that are peculiar to certain companies, for example, blizzards are peculiar to Dairy Queen, Rice Wraps are peculiar to Subway, and Timbits to Tim Hortons. There are also categories that contain similar menu items such as Chicken Strips, Chicken, Chicken McNuggets etc but are set into different categories by the different fast food restaurants.

3.0.3. Trans fat Content in Different Fast Food restaurants

We aim to assess the content of these unhealthy Trans fats across various restaurants; therefore, we will examine the trans fat content of menu items from these different establishments.

Figure 4

Trans fat Content by Restaurant



Note. This figure presents a bar chart depicting the trans fat content for different fast food restaurant chains. The x-axis shows the company names, while the y-axis displays the trans fat content in grams per serving.

The types of fats found in food include:

Unsaturated Fat: These are the good fat which doctors say should be the fat people eat most. They come as liquid oils instead of solid fats. On food packages, they are listed as mono or poly unsaturated fats.

Saturated Fat: These fats come from animals and are generally solid. They contain “bad” cholesterol and can contribute to heart disease. It is recommended that they make up less than 10% of daily caloric intake. Examples include whole-fat milk, butter, high-fat cheeses etc.

Trans Fat: These are the worst fats, and the FDA is forcing food companies to phase them out. They form when hydrogen is added to vegetable oil. This is done to increase shelf life but are bad for our health. Examples of food more likely to contain trans fat include stick margarine, pie crusts, ready-to-use frosting, some microwave popcorn and frozen pizza, etc.

From the visualization It looks like these companies are making efforts to phase out Trans Fat since their central tendencies are less than 0.5g. All of these companies have outliers with items that have more than 1g of trans fat.

McDonald's has the highest median trans fat content compared to the others. The median trans fat content for A&W, Diary Queen, and Subway is closest to 0g. Perhaps we can consider menu items from them as a "healthier" alternative to McDonald's and Tim Hortons.

3.0.4. Global F-Test

Anova to check the interaction model and Higher Order Model Against the Best Additive Model

$$H_0 : \beta_i = 0 \text{ vs } H_a : \beta_i \neq 0 \text{ (i = 1, 2, 3 ...)}$$

```
## Analysis of Variance Table
##
## Model 1: calories ~ 1
## Model 2: calories ~ company + calories_from_fat + total_fat_g + saturated_fat_g +
##           trans_fat_g + cholesterol_mg + sodium_mg + total_carbohydrates_g +
##           dietary_fiber_g + sugars_g + protein_g + total_fat_daily_value +
##           saturated_fat_daily_value + cholesterol_daily_value + sodium_daily_value +
##           total_carbohydrates_daily_value + dietary_fiber_daily_value +
##           protein_daily_value + vitamin_a_daily_value + vitamin_c_daily_value +
##           calcium_daily_value + iron_daily_value + category
##   Res.Df      RSS Df Sum of Sq    F          Pr(>F)
## 1    741 54112178
## 2    685 366434 56  53745744 1794.1 < 0.0000000000000022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3

Anova Table for Interaction Model and Additive Model

Source of Var.	Deg. of Freedom	Sum of Sqrs.	Mean Sqrs.	F-Stat
Regression	12	53695688	4474640.67	7832.164
Residual	729	416489	571.316	-
Total	741	54112178	-	-

Note. This table presents an ANOVA (Analysis of Variance) summary for the interaction model and additive model. It contains the Sources of Variation (Regression, Residual, and Total), the Degrees of Freedom and Sum of Squares associated with each Source of Variation, the Mean Squares and F-statistic.

With a p-value $< 0.0000000000000022 < 0.05$ we **reject** the null hypothesis H_0 at $\alpha = 0.05$. Therefore, at least one of the predictors is significant to our model.

3.0.5. VIF & Colinearity

The dependent variable is **Calories**. The predictor variables are the nutritional content, company name, and food category. These are listed in Section 2.1. We create a model with these variables and calculate the VIF to assess multicollinearity.

$$VIF(\hat{\beta}_j) = \frac{1}{1 - R_{X_j|X_{-j}}^2}$$

- VIFs=1 indicates that there is no collinearity between this independent variable and any others.

- $1 \leq VIFs \leq 5$ suggest that there is a moderate collinearity, but it is not severe enough to warrant corrective measures.
 - $VIFs > 5$ or 10 represent critical levels of multicollinearity.

```

## Call:
## imcdiag(mod = FullModel, method = "VIF")
##
## VIF Multicollinearity Diagnostics
##
##                                     VIF detection
## companydairy-queen                  Inf    1
## companymcdonalds                 4.1863   0
## companysubway                     16.1153   1
## companytim-hortons                5.5663   0
## calories_from_fat                 464.2926   1
## total_fat_g                       7281.1088   1
## saturated_fat_g                  46537.2040   1
## trans_fat_g                        4.7848   0
## cholesterol_mg                   7103.0287   1
## sodium_mg                          5204.5584   1
## total_carbohydrates_g             2517.5932   1
## dietary_fiber_g                  64951.6253   1
## sugars_g                           33.9754   1
## protein_g                          808538.2209   1
## total_fat_daily_value              6873.3671   1
## saturated_fat_daily_value         46640.2639   1
## cholesterol_daily_value            7093.7848   1
## sodium_daily_value                 5203.4538   1
## total_carbohydrates_daily_value   2487.7676   1
## dietary_fiber_daily_value          64937.9045   1
## protein_daily_value                808223.5832   1
## vitamin_a_daily_value               2.4068   0
## vitamin_c_daily_value               1.4552   0
## calcium_daily_value                 2.8013   0
## iron_daily_value                   4.0221   0
## categoryBaked Goods & Desserts   3.4744   0
## categoryBlizzards                  Inf    1
## categoryBreakfast                  10.1509   1
## categoryBurgers                    10.0997   1
## categoryBurgers and Sandwiches     Inf    1
## categoryChicken                   3.7133   0
## categoryChicken McNuggets          1.8310   0
## categoryChicken Strips              Inf    1
## categoryCoffee, Espresso & Tea    8.2491   0
## categoryCondiments                 2.3043   0
## categoryDesserts                  6.0328   0
## categoryDonuts                     4.1980   0
## categoryDrinks                     9.5041   0
## categoryHappy Meals                 2.1723   0
## categoryIce Cream Cones            Inf    1
## categoryMisty Slush                Inf    1
## categoryMooLattes                  Inf    1

```

```

## categoryNovelties           Inf    1
## categoryOrange Julius      Inf    1
## categoryRice Bowls          2.9755 0
## categoryRice Wraps          3.0448 0
## categorySalads              4.7096 0
## categorySandwiches & Wraps 5.2089 0
## categorySandwiches and Wraps 2.5873 0
## categoryShakes               Inf    1
## categorySides                4.0289 0
## categorySnacks & Sides       Inf    1
## categorySoups                2.1134 0
## categorySubway Sandwiches   15.5464 1
## categorySundaes & Royal Treats Inf    1
## categoryTimbits               3.3714 0
## categoryToppings and Spreads 1.8560 0
##
## Multicollinearity may be due to companydairy-queen companysubway calories_from_fat total_fat_g satur
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
##
## =====

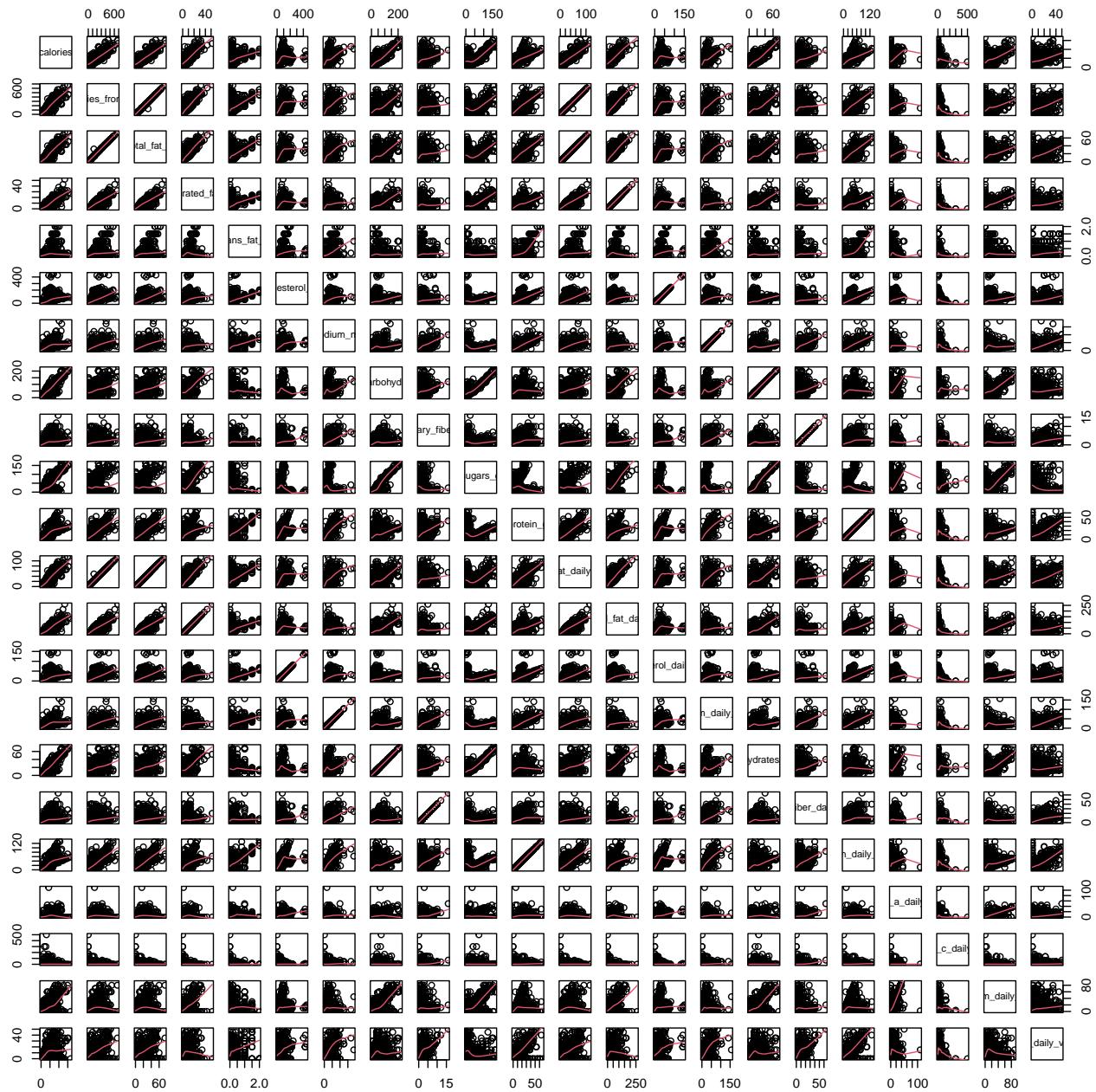
```

The VIF test detected collinearity. We will resolve this by removing predictors that are highly correlatable with one another. For the categorical predictors, there are certain food categories that are unique to only one company, e.g. Blizzards in Dairy Queen, Timbits in Tim Hortons etc. These can cause 1-1 relationships in some of the company sub-modules and category sub-modules causing collinearity between categories and companies. Since the categories are not standard across companies, we will remove the category column and keep the company column.

For the numeric predictors, we will use pair plots and correlation matrices to determine the predictors that are highly correlatable with each other (using a correlation cutoff of 0.8 or 80%) and select one either based on domain knowledge or the predictor that has the highest correlation with the dependent variable.

Figure 5

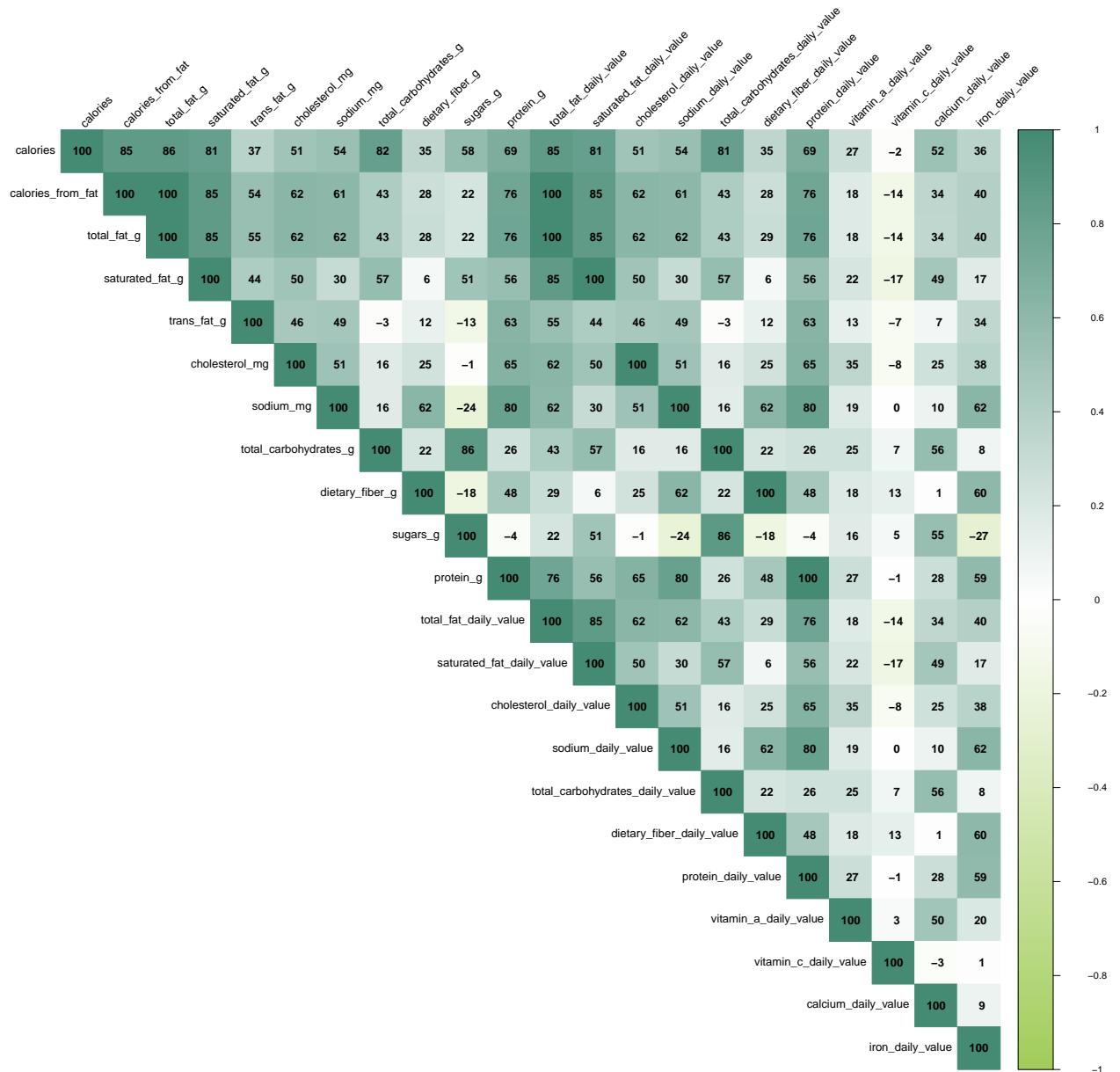
Pair Plot between Dependent and Predictor variable



Note. This figure presents a pair plot of the dependent calorie variable against all the nutrient predictors.

Figure 6

Correlation Matrix between Dependent and Predictor variable



Note. This figure presents a correlation Matrix between the dependent calorie variable against all the nutrient predictors.

- *Calories from Fat* is highly correlated with *total fat*, *saturated fat*, *total fat daily value*, and *saturated fat daily value*. Of these five predictors, we will use **total fat** which has the highest correlation with Calories among the others.
- *Cholesterol* and its *percentage daily value* have a 1-1 relationship. We will use the **cholesterol** nutrient value instead as it is a direct figure of the content of the menu item.
- *Sodium* is highly correlated with *protein*, *Sodium % daily value*, and *protein % daily value*. Of these, we will use **protein** which has the highest correlation with Calories among the others and is also a major food nutrient.
- *Total carbohydrates* is highly correlated with *sugars* and *Total carbohydrates % daily value*. Of these, we will use **Total carbohydrates** since it has the highest correlation with calories among the others and is a major food nutrient.

In **summary**, the predictors to be used for the full additive model are: **company**, **total_fat_g**,

`trans_fat_g`, `cholesterol_mg`, `total_carbohydrates_g`, `dietary_fiber_g`, `protein_g`, `vitamin_a_daily_value`, `vitamin_c_daily_value`, `calcium_daily_value`, and `iron_daily_value`.

3.0.6. Full Additive Model Analysis: Stepwise, Forward, Backward, All-Possible-Regressions, Individual T-test

We will use the predictors selected after the VIF test to build the first full model. First, we use the Stepwise, All-Possible-Regressions Selection procedures, and individual t-test to select a model.

- Using **Stepwise Regression Procedure**, `p_enter=0.05`, `p_remove=0.1`

```
##  
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),  
##      data = 1)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -67.599  -8.541  -1.376   5.240 283.051  
##  
## Coefficients:  
##              Estimate Std. Error t value     Pr(>|t|)  
## (Intercept) 8.90307  2.69305  3.306 0.000993 ***  
## total_fat_g 8.46865  0.11295 74.980 < 0.0000000000000002 ***  
## total_carbohydrates_g 4.04566  0.03363 120.301 < 0.0000000000000002 ***  
## protein_g    4.07397  0.15333 26.569 < 0.0000000000000002 ***  
## trans_fat_g  14.94406  3.68207  4.059 0.00005470 ***  
## iron_daily_value 0.45175  0.11358  3.977 0.00007666 ***  
## calcium_daily_value -0.15480  0.07025 -2.203 0.027878 *  
## dietary_fiber_g -1.52016  0.63953 -2.377 0.017712 *  
## companydairy-queen -17.58979  3.29026 -5.346 0.00000012 ***  
## companymcdonalds -8.90787  3.26108 -2.732 0.006456 **  
## companysubway    -9.06995  5.52172 -1.643 0.100898  
## companytim-hortons -9.48100  2.88770 -3.283 0.001075 **  
## vitamin_a_daily_value -0.23617  0.11437 -2.065 0.039286 *  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 23.9 on 729 degrees of freedom  
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9922  
## F-statistic:  7832 on 12 and 729 DF,  p-value: < 0.000000000000022
```

According to the stepwise regression procedure, the “best” set of predictors for calories are: `total_fat_g`, `total_carbohydrates_g`, `protein_g`, `trans_fat_g`, `iron_daily_value`, `calcium_daily_value`, `dietary_fiber_g`, `company`, and `vitamin_a_daily_value`.

- Using the **Forward Selection Regression Procedure**, `p_val = 0.05`

```
##  
## Call:  
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),  
##      data = 1)  
##
```

```

## Residuals:
##      Min     1Q   Median     3Q    Max
## -67.599 -8.541 -1.376  5.240 283.051
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.90307  2.69305  3.306  0.000993 ***
## total_fat_g            8.46865  0.11295 74.980 < 0.0000000000000002 ***
## total_carbohydrates_g  4.04566  0.03363 120.301 < 0.0000000000000002 ***
## protein_g               4.07397  0.15333 26.569 < 0.0000000000000002 ***
## trans_fat_g             14.94406 3.68207  4.059  0.00005470 ***
## iron_daily_value        0.45175  0.11358  3.977  0.00007666 ***
## calcium_daily_value     -0.15480  0.07025 -2.203  0.027878 *
## dietary_fiber_g         -1.52016  0.63953 -2.377  0.017712 *
## companydairy-queen     -17.58979 3.29026 -5.346  0.00000012 ***
## companymcdonalds       -8.90787  3.26108 -2.732  0.006456 **
## companysubway           -9.06995  5.52172 -1.643  0.100898
## companytim-hortons     -9.48100  2.88770 -3.283  0.001075 **
## vitamin_a_daily_value   -0.23617  0.11437 -2.065  0.039286 *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.9 on 729 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9922
## F-statistic:  7832 on 12 and 729 DF,  p-value: < 0.0000000000000022

```

According to the foward selection procedure, the “best” set of predictors for calories are: total_fat_g, total_carbohydrates_g, protein_g, trans_fat_g, iron_daily_value, calcium_daily_value, dietary_fiber_g, company, and vitamin_a_daily_value.

- Using the **Backward Elimination Regression Procedure**, p_val = 0.05

```

##
## Call:
## lm(formula = paste(response, "~", paste(c(include, cterms), collapse = " + ")),
##      data = 1)
##
## Residuals:
##      Min     1Q   Median     3Q    Max
## -67.599 -8.541 -1.376  5.240 283.051
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.90307  2.69305  3.306  0.000993 ***
## companydairy-queen     -17.58979 3.29026 -5.346  0.00000012 ***
## companymcdonalds       -8.90787  3.26108 -2.732  0.006456 **
## companysubway           -9.06995  5.52172 -1.643  0.100898
## companytim-hortons     -9.48100  2.88770 -3.283  0.001075 **
## total_fat_g             8.46865  0.11295 74.980 < 0.0000000000000002 ***
## trans_fat_g             14.94406 3.68207  4.059  0.00005470 ***
## total_carbohydrates_g  4.04566  0.03363 120.301 < 0.0000000000000002 ***
## dietary_fiber_g         -1.52016  0.63953 -2.377  0.017712 *
## protein_g               4.07397  0.15333 26.569 < 0.0000000000000002 ***
## vitamin_a_daily_value   -0.23617  0.11437 -2.065  0.039286 *

```

```

## calcium_daily_value   -0.15480   0.07025  -2.203          0.027878 *
## iron_daily_value      0.45175   0.11358   3.977          0.00007666 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.9 on 729 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9922
## F-statistic:  7832 on 12 and 729 DF,  p-value: < 0.0000000000000022

```

According to the backward elimination procedure, the “best” set of predictors for calories are: total_fat_g, total_carbohydrates_g, protein_g, trans_fat_g, iron_daily_value, calcium_daily_value, dietary_fiber_g, company, and vitamin_a_daily_value.

- Using the All-Possible-Regressions Selection Procedure (with AIC),

```

##      Predictors
## [1,] "total_fat_g"
## [2,] "total_fat_g total_carbohydrates_g"
## [3,] "total_fat_g total_carbohydrates_g protein_g"
## [4,] "company total_fat_g total_carbohydrates_g protein_g"
## [5,] "company total_fat_g trans_fat_g total_carbohydrates_g protein_g"
## [6,] "company total_fat_g trans_fat_g total_carbohydrates_g protein_g vitamin_a_daily_value"
## [7,] "company total_fat_g trans_fat_g total_carbohydrates_g protein_g vitamin_a_daily_value iron_da
## [8,] "company total_fat_g trans_fat_g total_carbohydrates_g dietary_fiber_g protein_g calcium_dai
## [9,] "company total_fat_g trans_fat_g total_carbohydrates_g dietary_fiber_g protein_g vitamin_a_dai
## [10,] "company total_fat_g trans_fat_g cholesterol_mg total_carbohydrates_g dietary_fiber_g protein_
## [11,] "company total_fat_g trans_fat_g cholesterol_mg total_carbohydrates_g dietary_fiber_g protein_
##       AdjR           Aic
## [1,] "0.730824753961338" "9445.24508352251"
## [2,] "0.974849538188439" "7687.34080794363"
## [3,] "0.991165014155613" "6912.08775615795"
## [4,] "0.991693002685027" "6870.33213744905"
## [5,] "0.991880915487392" "6854.34299054386"
## [6,] "0.99199668469669" "6844.67376264051"
## [7,] "0.99210378797623" "6835.66265393767"
## [8,] "0.99214154710029" "6833.09020919077"
## [9,] "0.992176525490477" "6830.76302060874"
## [10,] "0.992182223668979" "6831.20386127709"
## [11,] "0.992177058146535" "6832.67403749195"

```

Based on AIC, we need to pick the model subset with the smallest value of AIC. Using this criterion, the “best” predictors of Calories with the lowest AIC value of 6830.76302060874 are 9 variables: company, total_fat_g, trans_fat_g, total_carbohydrates_g, dietary_fiber_g, protein_g, vitamin_a_daily_value, calcium_daily_value, and iron_daily_value.

Based on adjusted R-squared, we need to pick the model subset with the highest value of adjusted R-squared. Using this criterion, the “best” predictors of Calories with the highest adjusted R^2 value of 0.992182223668979 are 10 variables: company, total_fat_g, trans_fat_g, cholesterol_mg, total_carbohydrates_g, dietary_fiber_g, protein_g, vitamin_a_daily_value, calcium_daily_value, and iron_daily_value.

- Using the All-Possible-Regressions Selection Procedure (with BIC)

```
## Subset selection object
```

```

## Call: regsubsets.formula(calories ~ company + total_fat_g + trans_fat_g +
##   cholesterol_mg + total_carbohydrates_g + dietary_fiber_g +
##   protein_g + vitamin_a_daily_value + vitamin_c_daily_value +
##   calcium_daily_value + iron_daily_value, data = mlr_df, nv = 11)
## 14 Variables (and intercept)
##          Forced in Forced out
## companydairy-queen      FALSE      FALSE
## companymcdonalds        FALSE      FALSE
## companysubway            FALSE      FALSE
## companytim-hortons      FALSE      FALSE
## total_fat_g              FALSE      FALSE
## trans_fat_g              FALSE      FALSE
## cholesterol_mg           FALSE      FALSE
## total_carbohydrates_g   FALSE      FALSE
## dietary_fiber_g          FALSE      FALSE
## protein_g                FALSE      FALSE
## vitamin_a_daily_value    FALSE      FALSE
## vitamin_c_daily_value    FALSE      FALSE
## calcium_daily_value      FALSE      FALSE
## iron_daily_value         FALSE      FALSE
## 1 subsets of each size up to 11
## Selection Algorithm: exhaustive
##          companydairy-queen companymcdonalds companysubway companytim-hortons
## 1  ( 1 )   " "          " "          " "          " "
## 2  ( 1 )   " "          " "          " "          " "
## 3  ( 1 )   " "          " "          " "          " "
## 4  ( 1 )   "*"          " "          " "          " "
## 5  ( 1 )   "*"          " "          " "          " "
## 6  ( 1 )   "*"          " "          " "          " "
## 7  ( 1 )   "*"          " "          " "          " "
## 8  ( 1 )   "*"          " "          " "          " "
## 9  ( 1 )   "*"          " "          " "          " "
## 10 ( 1 )   "*"          "*"          " "          "*"
## 11 ( 1 )   "*"          "*"          " "          "*"
##          total_fat_g trans_fat_g cholesterol_mg total_carbohydrates_g
## 1  ( 1 )   "*"          " "          " "          " "
## 2  ( 1 )   "*"          " "          " "          "*"
## 3  ( 1 )   "*"          " "          " "          "*"
## 4  ( 1 )   "*"          " "          " "          "*"
## 5  ( 1 )   "*"          "*"          " "          "*"
## 6  ( 1 )   "*"          "*"          " "          "*"
## 7  ( 1 )   "*"          "*"          " "          "*"
## 8  ( 1 )   "*"          "*"          " "          "*"
## 9  ( 1 )   "*"          "*"          " "          "*"
## 10 ( 1 )   "*"          "*"          " "          "*"
## 11 ( 1 )   "*"          "*"          " "          "*"
##          dietary_fiber_g protein_g vitamin_a_daily_value vitamin_c_daily_value
## 1  ( 1 )   " "          " "          " "          " "
## 2  ( 1 )   " "          " "          " "          " "
## 3  ( 1 )   " "          "*"          " "          " "
## 4  ( 1 )   " "          "*"          " "          " "
## 5  ( 1 )   " "          "*"          " "          " "
## 6  ( 1 )   " "          "*"          "*"          " "
## 7  ( 1 )   " "          "*"          "*"          " "

```

```

## 8  ( 1 )   "*"      "*"      " "      " "
## 9  ( 1 )   "*"      "*"      "*"      " "
## 10 ( 1 )  "*"      "*"      "*"      " "
## 11 ( 1 )  "*"      "*"      "*"      " "
##          calcium_daily_value iron_daily_value
## 1  ( 1 )   " "      " "      " "
## 2  ( 1 )   " "      " "      " "
## 3  ( 1 )   " "      " "      " "
## 4  ( 1 )   " "      " "      " "
## 5  ( 1 )   " "      " "      " "
## 6  ( 1 )   " "      " "      " "
## 7  ( 1 )   " "      " "      "*" 
## 8  ( 1 )   "*"      " "      " "
## 9  ( 1 )   "*"      " "      " "
## 10 ( 1 )  " "      " "      "*" 
## 11 ( 1 )  "*"      " "      " "

##          BIC
## [1,] -961.5787
## [2,] -2714.8736
## [3,] -3485.5173
## [4,] -3514.3110
## [5,] -3525.4977
## [6,] -3531.1323
## [7,] -3533.0196
## [8,] -3535.0676
## [9,] -3533.1767
## [10,] -3531.4591
## [11,] -3529.2261

```

Based on BIC, we need to pick the model subset with the smallest value of BIC. Using this criterion, the “best” predictors of Calories with the lowest BIC value of -3535.0676 are 8 variables: company, total_fat_g, trans_fat_g, total_carbohydrates_g, dietary_fiber_g, protein_g, calcium_daily_value, and iron_daily_value.

- Using the **Individual t-test** method,

The null hypothesis, H_o and the alternate hypothesis, H_a are as follows:

$$H_o : \beta_k = 0$$

$$H_a : \beta_k \neq 0 \quad (k = 1, 2, \dots, 11)$$

```

## 
## Call:
## lm(formula = calories ~ company + total_fat_g + trans_fat_g +
##      cholesterol_mg + total_carbohydrates_g + dietary_fiber_g +
##      protein_g + vitamin_a_daily_value + vitamin_c_daily_value +
##      calcium_daily_value + iron_daily_value, data = mlr_df)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -67.426  -8.418  -1.216   5.029 283.536
## 
## Coefficients:
##             Estimate Std. Error t value     Pr(>|t|)    
## company      1.2345   0.1234  9.8765  <2e-16 ***
## total_fat_g -0.5678   0.0567 -10.0123  <2e-16 ***
## trans_fat_g  0.8912   0.0891  10.0123  <2e-16 ***
## cholesterol_mg  0.0012   0.0012  1.0000    0.3162    
## total_carbohydrates_g  0.0023   0.0023  1.0000    0.3162    
## dietary_fiber_g  0.0034   0.0034  1.0000    0.3162    
## protein_g     0.0045   0.0045  1.0000    0.3162    
## vitamin_a_daily_value  0.0056   0.0056  1.0000    0.3162    
## vitamin_c_daily_value  0.0067   0.0067  1.0000    0.3162    
## calcium_daily_value  0.0078   0.0078  1.0000    0.3162    
## iron_daily_value  0.0089   0.0089  1.0000    0.3162    
## 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
```

```

## (Intercept) 8.41789 2.74631 3.065 0.00226 **
## companydairy-queen -17.22948 3.30256 -5.217 0.000000237 ***
## companymcdonalds -8.79322 3.27581 -2.684 0.00743 **
## companysubway -7.67218 5.60992 -1.368 0.17186
## companytim-hortons -9.14876 2.92511 -3.128 0.00183 **
## total_fat_g 8.45007 0.12058 70.081 < 0.0000000000000002 ***
## trans_fat_g 15.02473 3.68519 4.077 0.000050642 ***
## cholesterol_mg 0.02677 0.02126 1.259 0.20844
## total_carbohydrates_g 4.04646 0.03417 118.431 < 0.0000000000000002 ***
## dietary_fiber_g -1.59532 0.64470 -2.475 0.01357 *
## protein_g 4.00989 0.15993 25.072 < 0.0000000000000002 ***
## vitamin_a_daily_value -0.27515 0.11822 -2.327 0.02022 *
## vitamin_c_daily_value 0.02082 0.02889 0.721 0.47137
## calcium_daily_value -0.14970 0.07042 -2.126 0.03385 *
## iron_daily_value 0.45734 0.11366 4.024 0.000063250 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.9 on 727 degrees of freedom
## Multiple R-squared: 0.9923, Adjusted R-squared: 0.9922
## F-statistic: 6714 on 14 and 727 DF, p-value: < 0.0000000000000022

```

Based on the t-test with a significance level of 0.05, the “best” predictors of Calories are company, total_fat_g, trans_fat_g, total_carbohydrates_g, dietary_fiber_g, protein_g, vitamin_a_daily_value, calcium_daily_value, and iron_daily_value.

Comparing the results, there are 9 independent variables that are consistently selected as the “best” predictors confirmed by the Individual T-Test, Stepwise Regression, Forward Selection, Backward Selection and AIC. They are **company**, **total_fat_g**, **trans_fat_g**, **total_carbohydrates_g**, **dietary_fiber_g**, **protein_g**, **vitamin_a_daily_value**, **calcium_daily_value**, and **iron_daily_value**.

- The final **best additive model** is shown below:

```

##
## Call:
## lm(formula = calories ~ total_fat_g + total_carbohydrates_g +
##     protein_g + trans_fat_g + iron_daily_value + calcium_daily_value +
##     dietary_fiber_g + company + vitamin_a_daily_value, data = mlr_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -67.599  -8.541  -1.376   5.240 283.051
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.90307 2.69305 3.306 0.000993 ***
## total_fat_g 8.46865 0.11295 74.980 < 0.0000000000000002 ***
## total_carbohydrates_g 4.04566 0.03363 120.301 < 0.0000000000000002 ***
## protein_g 4.07397 0.15333 26.569 < 0.0000000000000002 ***
## trans_fat_g 14.94406 3.68207 4.059 0.00005470 ***
## iron_daily_value 0.45175 0.11358 3.977 0.00007666 ***
## calcium_daily_value -0.15480 0.07025 -2.203 0.027878 *
## dietary_fiber_g -1.52016 0.63953 -2.377 0.017712 *
## companydairy-queen -17.58979 3.29026 -5.346 0.00000012 ***

```

```

## companymcdonalds      -8.90787   3.26108  -2.732      0.006456 **
## companysubway        -9.06995   5.52172  -1.643      0.100898
## companytim-hortons   -9.48100   2.88770  -3.283      0.001075 **
## vitamin_a_daily_value -0.23617   0.11437  -2.065      0.039286 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.9 on 729 degrees of freedom
## Multiple R-squared:  0.9923, Adjusted R-squared:  0.9922
## F-statistic:  7832 on 12 and 729 DF,  p-value: < 0.0000000000000022

```

3.0.7. Interaction Model

To improve the first order model, we will build an interaction model on the additive model and perform an individual t-test (with significance level of 0.05) on the independent predictors and interaction terms.

$$H_o : \beta_i = 0$$

$$H_a : \beta_i \neq 0 \text{ (i=1,2,...,p)}$$

where p is the number of independent variables and interaction terms in the regression model.

```

##
## Call:
## lm(formula = calories ~ total_fat_g + total_carbohydrates_g +
##     protein_g + trans_fat_g + iron_daily_value + calcium_daily_value +
##     dietary_fiber_g + company + vitamin_a_daily_value + total_fat_g:total_carbohydrates_g +
##     total_fat_g:calcium_daily_value + total_fat_g:company + total_fat_g:vitamin_a_daily_value +
##     total_carbohydrates_g:trans_fat_g + total_carbohydrates_g:iron_daily_value +
##     total_carbohydrates_g:calcium_daily_value + total_carbohydrates_g:vitamin_a_daily_value +
##     protein_g:calcium_daily_value + trans_fat_g:iron_daily_value +
##     trans_fat_g:vitamin_a_daily_value + iron_daily_value:dietary_fiber_g +
##     iron_daily_value:company + calcium_daily_value:dietary_fiber_g +
##     calcium_daily_value:company + dietary_fiber_g:company, data = mlr_df)
##
## Residuals:
##      Min        1Q        Median       3Q        Max
## -115.836    -7.452    -0.781     5.529    278.489
##
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)                13.706172  3.594496  3.813
## total_fat_g                  8.217141  0.273643 30.029
## total_carbohydrates_g       3.804626  0.053249 71.450
## protein_g                   4.054013  0.192179 21.095
## trans_fat_g                 46.020517  7.175237  6.414
## iron_daily_value             -0.066106  0.306124 -0.216
## calcium_daily_value          -0.371490  0.208480 -1.782
## dietary_fiber_g              3.989180  1.638344  2.435
## companydairy-queen          -6.128804  4.869696 -1.259
## companymcdonalds            -16.084629  5.797317 -2.774
## companysubway                 -55.032091 12.929542 -4.256
## companytim-hortons           -12.393748  3.986586 -3.109
## vitamin_a_daily_value         0.029895  0.258384  0.116

```

## total_fat_g:total_carbohydrates_g	0.007618	0.001905	4.000
## total_fat_g:calcium_daily_value	-0.038291	0.007332	-5.223
## total_fat_g:companydairy-queen	-0.633735	0.308290	-2.056
## total_fat_g:companymcdonalds	-0.001811	0.327113	-0.006
## total_fat_g:companysubway	-0.246301	0.396639	-0.621
## total_fat_g:companytim-hortons	0.236509	0.313130	0.755
## total_fat_g:vitamin_a_daily_value	0.049885	0.012388	4.027
## total_carbohydrates_g:trans_fat_g	-0.499949	0.105892	-4.721
## total_carbohydrates_g:iron_daily_value	0.017269	0.003143	5.495
## total_carbohydrates_g:calcium_daily_value	0.007894	0.002027	3.894
## total_carbohydrates_g:vitamin_a_daily_value	-0.017592	0.003378	-5.208
## protein_g:calcium_daily_value	0.025840	0.009242	2.796
## trans_fat_g:iron_daily_value	-0.591765	0.198757	-2.977
## trans_fat_g:vitamin_a_daily_value	1.223408	0.442610	2.764
## iron_daily_value:dietary_fiber_g	-0.295408	0.051121	-5.779
## iron_daily_value:companydairy-queen	1.774618	0.423102	4.194
## iron_daily_value:companymcdonalds	-0.274011	0.361706	-0.758
## iron_daily_value:companysubway	0.677258	0.660323	1.026
## iron_daily_value:companytim-hortons	0.018305	0.344934	0.053
## calcium_daily_value:dietary_fiber_g	-0.168007	0.034729	-4.838
## calcium_daily_value:companydairy-queen	-0.013517	0.185320	-0.073
## calcium_daily_value:companymcdonalds	0.654328	0.195643	3.344
## calcium_daily_value:companysubway	-0.304285	1.043273	-0.292
## calcium_daily_value:companytim-hortons	0.343876	0.201202	1.709
## dietary_fiber_g:companydairy-queen	-4.858326	1.891707	-2.568
## dietary_fiber_g:companymcdonalds	1.899301	1.731539	1.097
## dietary_fiber_g:companysubway	7.593078	2.969514	2.557
## dietary_fiber_g:companytim-hortons	-1.575298	2.177562	-0.723
##		Pr(> t)	
## (Intercept)		0.000149	***
## total_fat_g	< 0.0000000000000002		***
## total_carbohydrates_g	< 0.0000000000000002		***
## protein_g	< 0.0000000000000002		***
## trans_fat_g	0.000000000261		***
## iron_daily_value	0.829093		
## calcium_daily_value	0.075200		.
## dietary_fiber_g	0.015145		*
## companydairy-queen	0.208608		
## companymcdonalds	0.005676		**
## companysubway	0.000023611272		***
## companytim-hortons	0.001954		**
## vitamin_a_daily_value	0.907923		
## total_fat_g:total_carbohydrates_g	0.000070031531		***
## total_fat_g:calcium_daily_value	0.000000232608		***
## total_fat_g:companydairy-queen	0.040186		*
## total_fat_g:companymcdonalds	0.995585		
## total_fat_g:companysubway	0.534821		
## total_fat_g:companytim-hortons	0.450319		
## total_fat_g:vitamin_a_daily_value	0.000062725857		***
## total_carbohydrates_g:trans_fat_g	0.000002830805		***
## total_carbohydrates_g:iron_daily_value	0.00000054702		***
## total_carbohydrates_g:calcium_daily_value	0.000108		***
## total_carbohydrates_g:vitamin_a_daily_value	0.000000250838		***
## protein_g:calcium_daily_value	0.005314		**

```

## trans_fat_g:iron_daily_value          0.003008 **
## trans_fat_g:vitamin_a_daily_value    0.005858 **
## iron_daily_value:dietary_fiber_g    0.000000011332 ***
## iron_daily_value:companydairy-queen 0.000030883119 ***
## iron_daily_value:companymcdonalds   0.448975
## iron_daily_value:companysubway     0.305412
## iron_daily_value:companytim-hortons 0.957692
## calcium_daily_value:dietary_fiber_g 0.000001616523 ***
## calcium_daily_value:companydairy-queen 0.941877
## calcium_daily_value:companymcdonalds 0.000868 ***
## calcium_daily_value:companysubway   0.770630
## calcium_daily_value:companytim-hortons 0.087873 .
## dietary_fiber_g:companydairy-queen 0.010428 *
## dietary_fiber_g:companymcdonalds   0.273068
## dietary_fiber_g:companysubway     0.010767 *
## dietary_fiber_g:companytim-hortons 0.469662
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.38 on 701 degrees of freedom
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.9943
## F-statistic:  3238 on 40 and 701 DF,  p-value: < 0.0000000000000022

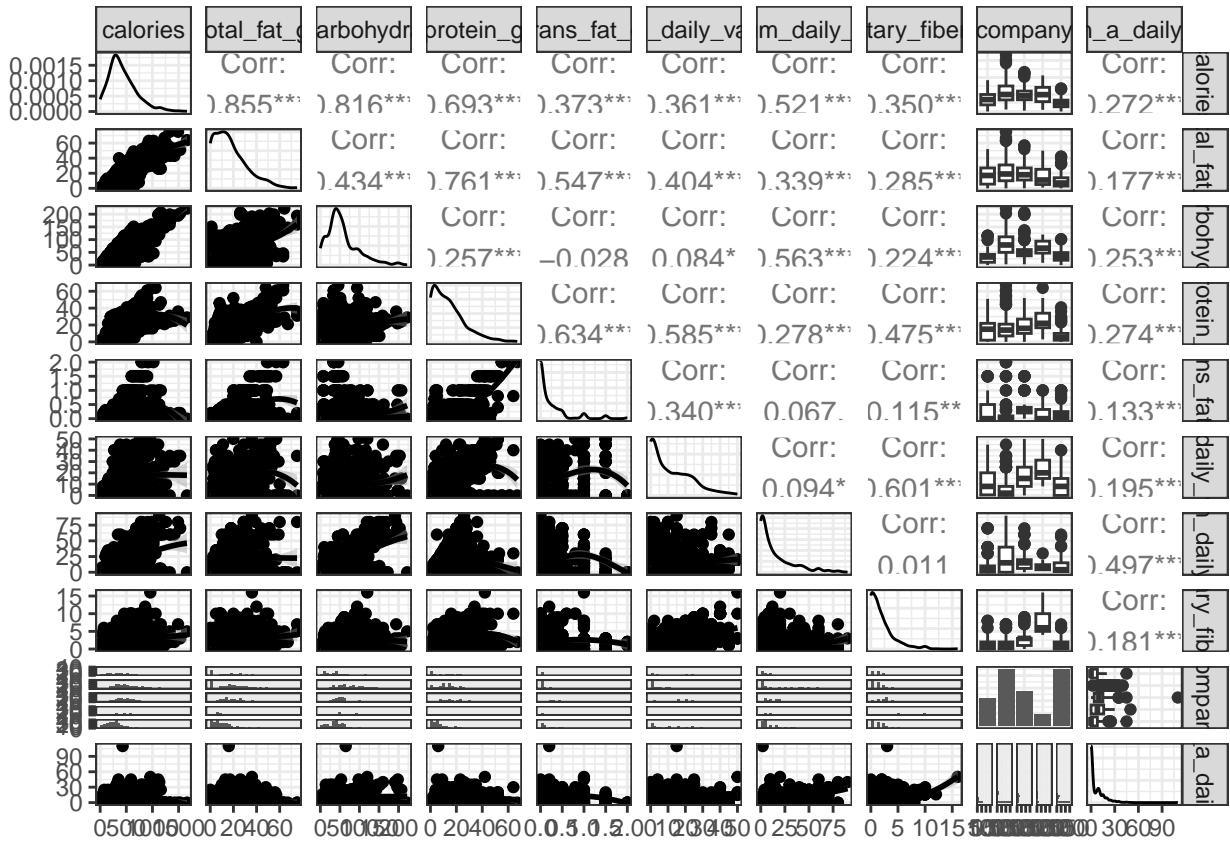
```

3.0.8. Higher Order Model

For further model improvement procedures, we check for the effect of higher-order terms in the model, starting from the predictor with highest correlation to the dependent variable, **Calories**.

Figure 7

Pair Plot between Dependent and Predictor variable for Higher Order



```

## 
## Call:
## lm(formula = calories ~ total_fat_g + total_carbohydrates_g +
##     I(total_carbohydrates_g^2) + protein_g + trans_fat_g + iron_daily_value +
##     calcium_daily_value + dietary_fiber_g + company + vitamin_a_daily_value +
##     total_fat_g:calcium_daily_value + total_fat_g:vitamin_a_daily_value +
##     total_carbohydrates_g:trans_fat_g + total_carbohydrates_g:iron_daily_value +
##     total_carbohydrates_g:calcium_daily_value + total_carbohydrates_g:vitamin_a_daily_value +
##     protein_g:calcium_daily_value + trans_fat_g:iron_daily_value +
##     trans_fat_g:vitamin_a_daily_value + iron_daily_value:dietary_fiber_g +
##     iron_daily_value:company + calcium_daily_value:dietary_fiber_g +
##     calcium_daily_value:company + dietary_fiber_g:company, data = mlr_df)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -118.400  -7.219  -1.040   5.602  277.904 
## 
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)                14.4348711  3.4478028  4.187
## total_fat_g                  8.4292050  0.1384362 60.889
## total_carbohydrates_g        3.6590814  0.0738427 49.552
## I(total_carbohydrates_g^2)    0.0020975  0.0004906  4.275
## protein_g                     3.8932770  0.1802769 21.596
## trans_fat_g                   38.5201042  6.2805028  6.133
## iron_daily_value              -0.0180614  0.2430349 -0.074
## calcium_daily_value           -0.1689378  0.2171067 -0.778

```

## dietary_fiber_g	5.1892001	1.5838277	3.276
## companydairy-queen	-10.4594706	4.4887114	-2.330
## companymcdonalds	-12.4284044	5.5847870	-2.225
## companysubway	-54.2268117	12.8214355	-4.229
## companytim-hortons	-9.6559355	3.7502648	-2.575
## vitamin_a_daily_value	0.0174021	0.2485788	0.070
## total_fat_g:calcium_daily_value	-0.0318614	0.0072959	-4.367
## total_fat_g:vitamin_a_daily_value	0.0551600	0.0113005	4.881
## total_carbohydrates_g:trans_fat_g	-0.4205755	0.0987503	-4.259
## total_carbohydrates_g:iron_daily_value	0.0178821	0.0031037	5.762
## total_carbohydrates_g:calcium_daily_value	0.0051390	0.0021671	2.371
## total_carbohydrates_g:vitamin_a_daily_value	-0.0179562	0.0033122	-5.421
## protein_g:calcium_daily_value	0.0227715	0.0092412	2.464
## trans_fat_g:iron_daily_value	-0.3884638	0.1837424	-2.114
## trans_fat_g:vitamin_a_daily_value	1.0490068	0.4366089	2.403
## iron_daily_value:dietary_fiber_g	-0.3029929	0.0499671	-6.064
## iron_daily_value:companydairy-queen	1.5855166	0.3636769	4.360
## iron_daily_value:companymcdonalds	-0.3054496	0.2495350	-1.224
## iron_daily_value:companysubway	0.6390379	0.6307462	1.013
## iron_daily_value:companytim-hortons	0.0225782	0.2822477	0.080
## calcium_daily_value:dietary_fiber_g	-0.1575161	0.0343815	-4.581
## calcium_daily_value:companydairy-queen	-0.1074881	0.1780362	-0.604
## calcium_daily_value:companymcdonalds	0.5402823	0.1916495	2.819
## calcium_daily_value:companysubway	-0.2015921	0.9704487	-0.208
## calcium_daily_value:companytim-hortons	0.3205885	0.1939641	1.653
## dietary_fiber_g:companydairy-queen	-6.4392831	1.8497889	-3.481
## dietary_fiber_g:companymcdonalds	1.2133270	1.7048265	0.712
## dietary_fiber_g:companysubway	6.3959908	2.8820572	2.219
## dietary_fiber_g:companytim-hortons	-1.9378070	2.1317720	-0.909
##	Pr(> t)		
## (Intercept)	0.00003188946	***	
## total_fat_g	< 0.0000000000000002	***	
## total_carbohydrates_g	< 0.0000000000000002	***	
## I(total_carbohydrates_g^2)	0.00002174432	***	
## protein_g	< 0.0000000000000002	***	
## trans_fat_g	0.00000000143	***	
## iron_daily_value	0.94078		
## calcium_daily_value	0.43675		
## dietary_fiber_g	0.00110	**	
## companydairy-queen	0.02008	*	
## companymcdonalds	0.02637	*	
## companysubway	0.00002652287	***	
## companytim-hortons	0.01023	*	
## vitamin_a_daily_value	0.94421		
## total_fat_g:calcium_daily_value	0.00001448289	***	
## total_fat_g:vitamin_a_daily_value	0.00000130517	***	
## total_carbohydrates_g:trans_fat_g	0.00002332091	***	
## total_carbohydrates_g:iron_daily_value	0.00000001245	***	
## total_carbohydrates_g:calcium_daily_value	0.01799	*	
## total_carbohydrates_g:vitamin_a_daily_value	0.00000008138	***	
## protein_g:calcium_daily_value	0.01397	*	
## trans_fat_g:iron_daily_value	0.03485	*	
## trans_fat_g:vitamin_a_daily_value	0.01654	*	
## iron_daily_value:dietary_fiber_g	0.00000000217	***	

```

## iron_daily_value:companydairy-queen          0.00001496589 ***
## iron_daily_value:companymcdonalds           0.22133
## iron_daily_value:companysubway             0.31134
## iron_daily_value:companytim-hortons        0.93626
## calcium_daily_value:dietary_fiber_g         0.00000546114 ***
## calcium_daily_value:companydairy-queen       0.54621
## calcium_daily_value:companymcdonalds        0.00495 **
## calcium_daily_value:companysubway           0.83550
## calcium_daily_value:companytim-hortons       0.09881 .
## dietary_fiber_g:companydairy-queen          0.00053 ***
## dietary_fiber_g:companymcdonalds            0.47689
## dietary_fiber_g:companysubway              0.02679 *
## dietary_fiber_g:companytim-hortons          0.36365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.34 on 705 degrees of freedom
## Multiple R-squared:  0.9946, Adjusted R-squared:  0.9943
## F-statistic:  3613 on 36 and 705 DF,  p-value: < 0.0000000000000022

```

The step-by-step process from additive to higher order model building can be found in the appendix for reproducibility.

3.0.9. Model Selection

Comparing the residual standard error (RSE) and Adjusted R-squared of the valid additive, interaction, and higher-order models respectively:

1. Additive Model: RSE is 23.9, Adjusted R-squared is 0.9922.
2. Interaction Model: RSE is 20.38, Adjusted R-squared is 0.9943.
3. Higher Order Model: RSE is 20.34, Adjusted R-squared is 0.9943.

While the interaction and higher-order models reduced the RSE by more than 3 points, the adjusted R-squared value changed very slightly.

In order to avoid overfitting as the adjusted R-squared is almost perfect (~99%), we will be selecting the additive model as the model for predictions.

With the additive model:

- 99.22% of the variation in Calories is explained by the regression model which depends on the total fat (g), total carbohydrates (g), protein (g), trans fat (g), iron % daily value, calcium % daily value, dietary fiber (g), company, and vitamin A % daily value.
- the average difference between the fitted values from the regression model and the actual observed values is 23.9 calories.

3.0.10. Validation of Model Assumptions

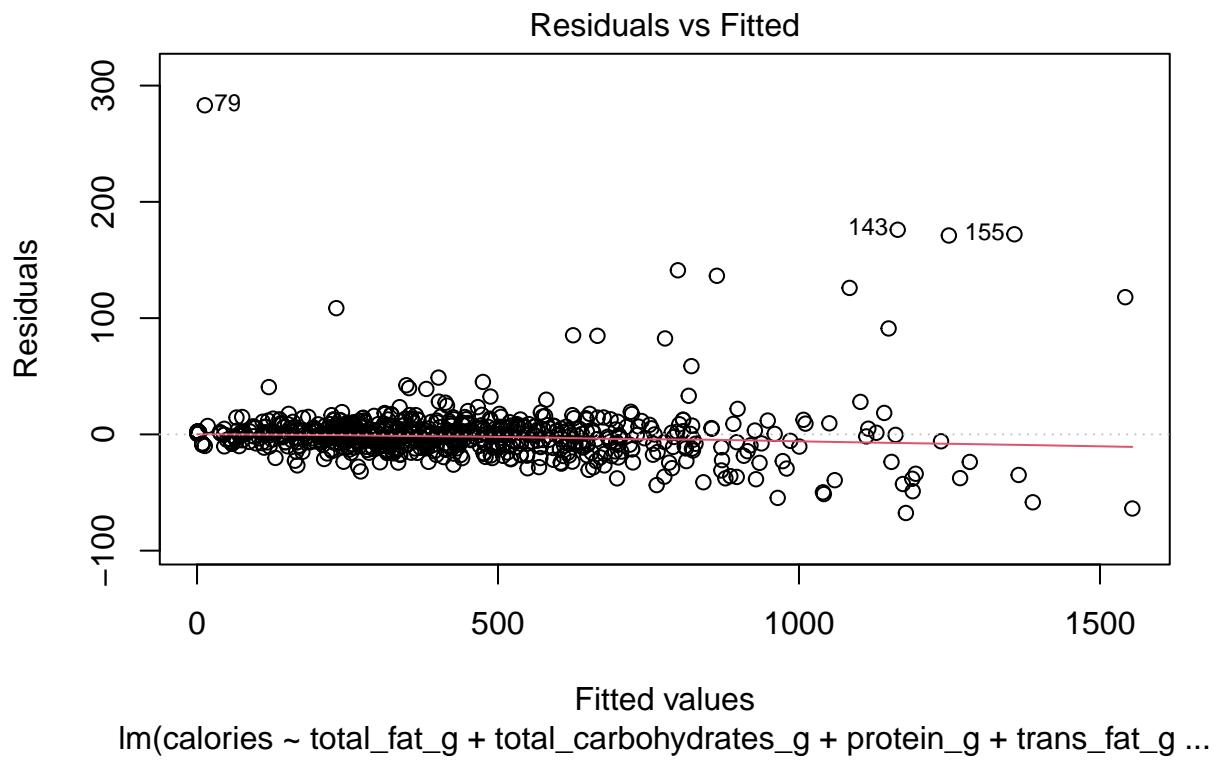
We will validate the additive model for linearity, equal variance, and normality.

Linearity Assumption

Plotting the residual vs fitted values plot,

Figure 8

Residual vs Fitted values



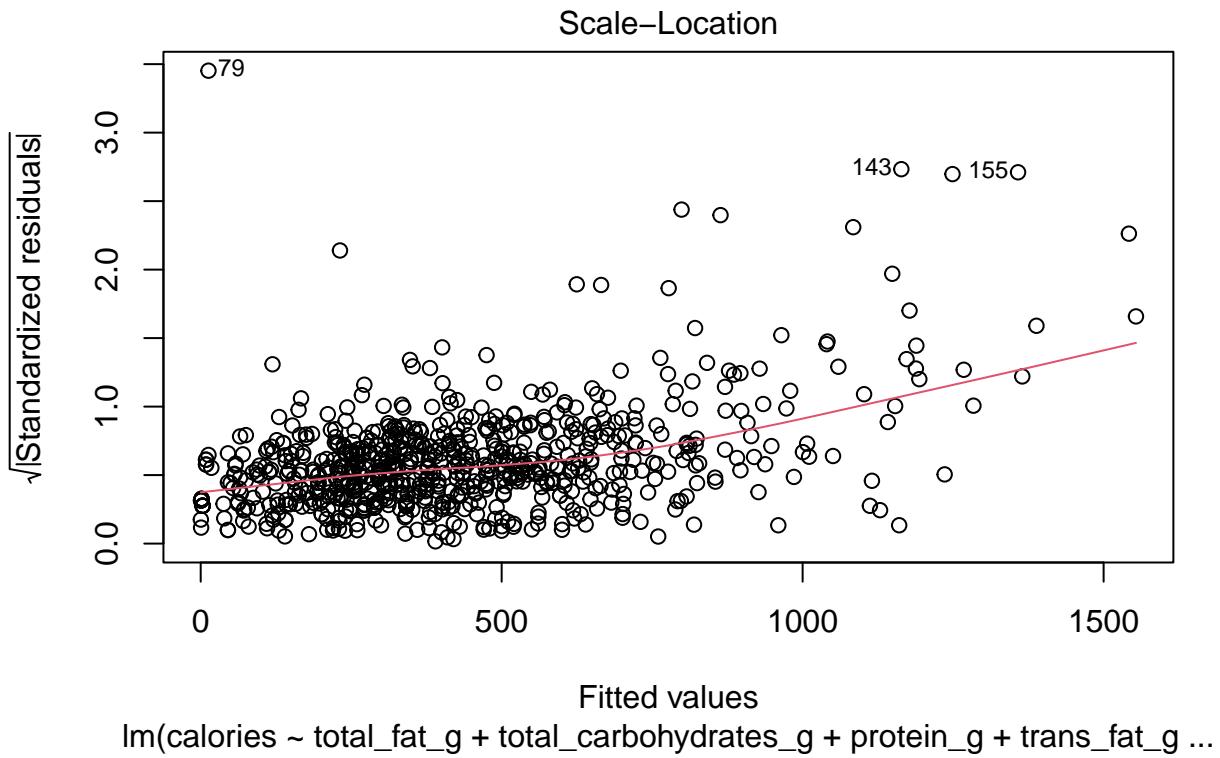
The model follows a fairly normal line and passes the assumption of linearity.

Equal Variance Assumption

Visualizing the scale-location plot,

Figure 9

Scale-Location Plot



Conducting the Breusch- Pagan Test to test the following hypothesis:

H_0 : heteroscedasticity is not present (homoscedasticity)

H_a : heteroscedasticity is present

or

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$

$H_0 : \text{at least one } \sigma_i^2 \text{ is different from the others } i = 1, 2, \dots, n$

```
##  
## studentized Breusch-Pagan test  
##  
## data: caloriesAdditiveModel  
## BP = 57.131, df = 12, p-value = 0.00000007484
```

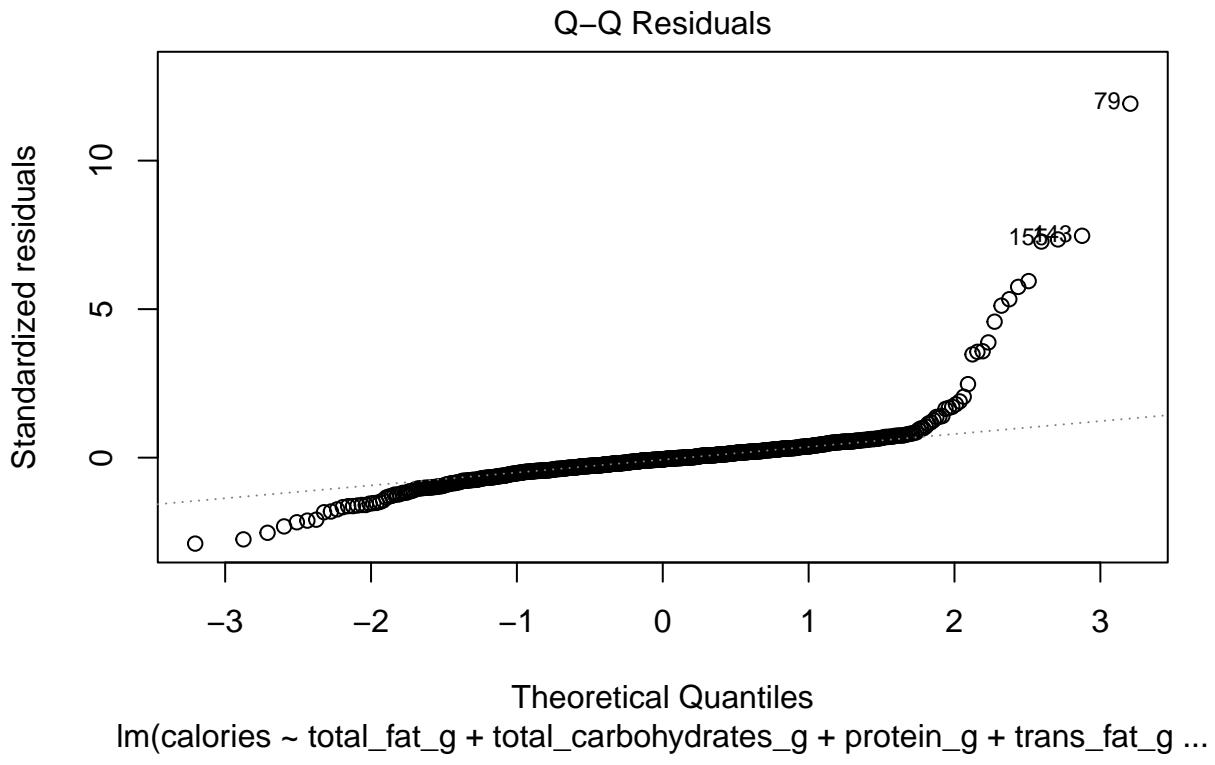
At a significance level of 0.05, we reject the null hypothesis and conclude that heteroscedasticity is present.

Normality Assumption

From the Q-Q plot,

Figure 10

Q-Q Plot



Using the Shapiro-Wilk test (S-W):

H_0 : the sample data are significantly normally distributed

H_a : the sample data are not significantly normally distributed

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(caloriesAdditiveModel)  
## W = 0.60954, p-value < 0.000000000000022
```

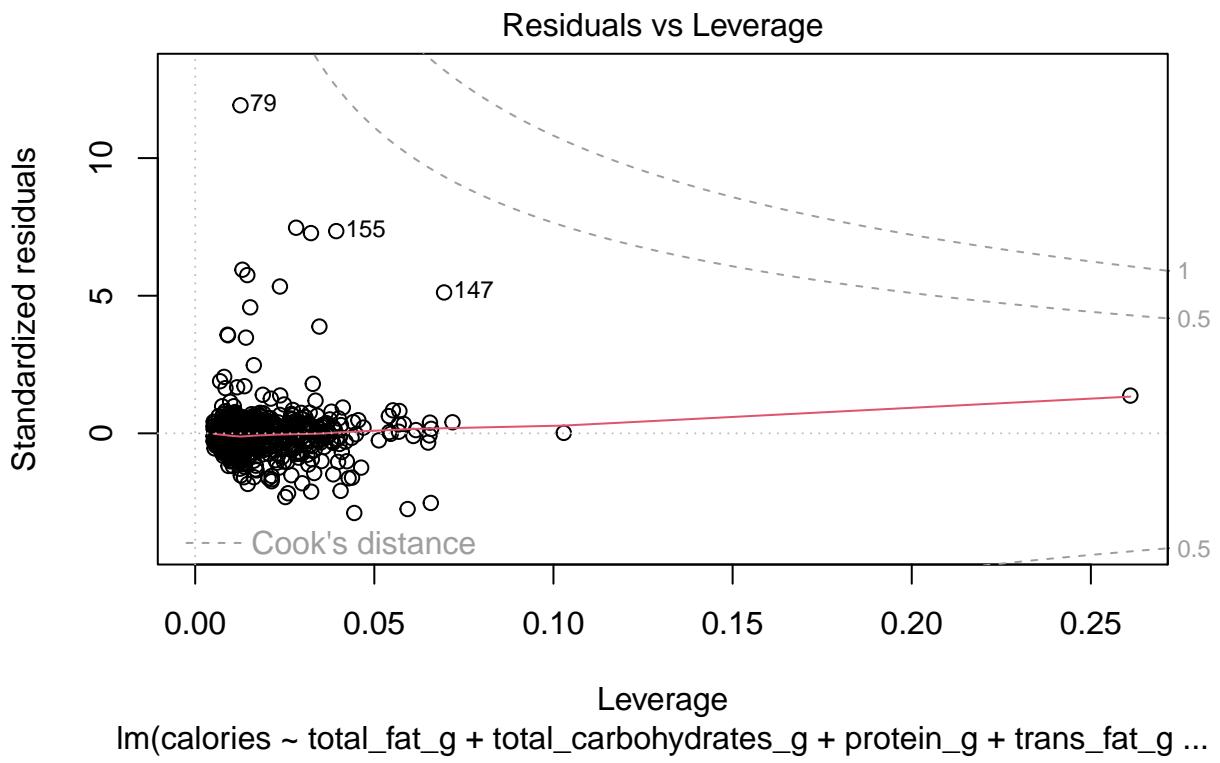
At a significance level of 0.05, we reject the null hypothesis that we have normality.

Outliers

We also check to see whether there are any significant outliers in the model.

Figure 11

Cooks Distance Plot to Identify Outliers



All standardized residuals in the model lie within the Cook's distance less than 0.5 so we can conclude that there are no outliers that will significantly change the output of the model.

3.1. Results

1. Menu items from Tim Hortons have the lowest calorie distribution compared to A&W, McDonalds, Diary Queen, and Subway.
 2. The median trans fat content for A&W, Diary Queen, and Subway is closest to 0g. Perhaps we can consider menu items from them as a “healthier” alternative to McDonald’s and Tim Hortons.
 3. Calories can be explained by the regression model which depends on the total fat (g), total carbohydrates (g), protein (g), trans fat (g), iron % daily value, calcium % daily value, dietary fiber (g), company, and vitamin A % daily value.
- The equation is as follows:

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 17.58979 \cdot \text{companydairy_queen} \\ - 8.90787 \cdot \text{companymcdonalds} - 9.06995 \cdot \text{companysubway} \\ - 9.48100 \cdot \text{companytim_hortons} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

The Submodels are given below

For Company A&W

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

For Company Dairy Queen

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 17.58979 \cdot \text{companydairy_queen} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

For Company Mcdonalds

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 8.90787 \cdot \text{companymcdonalds} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

For company Subway

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 9.06995 \cdot \text{companysubway} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

For Company Tim-Horton

$$\hat{\text{calories}} = 8.90307 + 8.46865 \cdot \text{total_fat_g} + 4.04566 \cdot \text{total_carbohydrates_g} + 4.07397 \cdot \text{protein_g} \\ + 14.94406 \cdot \text{trans_fat_g} + 0.45175 \cdot \text{iron_daily_value} - 0.15480 \cdot \text{calcium_daily_value} \\ - 1.52016 \cdot \text{dietary_fiber_g} - 9.48100 \cdot \text{companym_hortons} - 0.23617 \cdot \text{vitamin_a_daily_value}$$

- The interpretation is as follows:

1. **8.90307** represents the baseline calories when all other predictors (e.g., total fat, protein) are zero.
2. For every additional gram of total fat by keeping all other predictors constant, the calorie count **increases** by **8.47** calories.
3. For every additional gram of carbohydrate by keeping all other predictors constant, the calorie count **increases** by **4.05** calories.
4. For every additional gram of protein by keeping all other predictors constant, the calorie count **increases** by **4.07** calories.
5. For every additional gram of trans fat by keeping all other predictors constant, the calorie count **increases** by **14.94** calories.
6. For every 1% increase in the daily value of iron by keeping all other predictors constant, the calorie count **increases** by **0.45** calories.
7. For every 1% increase in the daily value of calcium by keeping all other predictors constant, the calorie count **decreases** by **0.15** calories.
8. For every additional gram of dietary fiber by keeping all other predictors constant, the calorie count **decreases** by **1.52** calories.
9. For every 1% increase in the daily value of Vitamin A by keeping all other predictors constant, the calorie count **decreases** by **0.24** calories.
10. Foods from Dairy Queen tend to have **17.59** fewer calories compared to the A&W by keeping all other predictors constant.
11. Foods from McDonald's have **8.91** fewer calories compared to the A&W by keeping all other predictors constant.
12. Foods from Subway have **9.07** fewer calories compared to the A&W by keeping all other predictors constant.
13. Foods from Tim Hortons have **9.48** fewer calories compared to the A&W by keeping all other predictors constant.

3.2. Model Prediction and Confidence Interval

3.2.1. Confidence Interval

```
##                      2.5 %      97.5 %
## (Intercept)      3.6159994 14.19013313
## total_fat_g       8.2469107  8.69038738
## total_carbohydrates_g 3.9796420  4.11168641
## protein_g         3.7729403  4.37500197
## trans_fat_g        7.7153253 22.17278728
## iron_daily_value   0.2287623  0.67473878
## calcium_daily_value -0.2927252 -0.01687393
```

```

## dietary_fiber_g      -2.7757049  -0.26460728
## companydairy-queen   -24.0493052  -11.13027540
## companymcdonalds    -15.3100887  -2.50564767
## companysubway        -19.9103133   1.77041837
## companytim-hortons   -15.1501921  -3.81180978
## vitamin_a_daily_value -0.4607042  -0.01162851

```

Sample: Dietary Fiber, Total Carbohydrate:

- The 95% confidence interval for regression coefficient for dietary_fiber_g is (-0.26460728, -0.26460728). This means that we conclude 95% confidence that for every **1 gram** increase in dietary_fiber_g, the calories **decreases** between **0.26460728** and **0.26460728** calories while holding the total_fat_g, total_carbohydrates_g, protein_g, trans_fat_g, iron_daily_value, calcium_daily_value, company and vitamin_a_daily_value constant.
- The 95% confidence interval for regression coefficient for total_carbohydrates_g is (-3.9796420, -4.11168641). This means that we conclude 95% confidence that for every **1 gram** increase in total_carbohydrates_g, the calories **increases** between **3.9796420** and **4.11168641** calories while holding the total_fat_g, protein_g, trans_fat_g, iron_daily_value, calcium_daily_value, dietary_fiber_g, company and vitamin_a_daily_value constant.

3.2.2. Prediction

- We can get a range of the values for prediction:

```

## [1] 0 75
## [1] 0 222
## [1] 0 67
## [1] 0 2
## [1] 0 50
## [1] 0 90
## [1] 0 16
## [1] "a-w"      "tim-hortons"
## [1] 0 110

```

- or alternatively we can select a particular row in the data to test the models accuracy:

```

##   company calories calories_from_fat total_fat_g saturated_fat_g trans_fat_g
## 32     a-w       620            351         39          13           1
##   cholesterol_mg sodium_mg total_carbohydrates_g dietary_fiber_g sugars_g
## 32            80        800            37           1           9
##   protein_g total_fat_daily_value saturated_fat_daily_value
## 32          27            60            65
##   cholesterol_daily_value sodium_daily_value total_carbohydrates_daily_value
## 32                  27            33             12
##   dietary_fiber_daily_value protein_daily_value vitamin_a_daily_value
## 32                      4            54             8
##   vitamin_c_daily_value calcium_daily_value iron_daily_value category
## 32                      8            20            25  Burgers

```

For the Single Mozza Burger menu item from A&W with these content values - Total Fat = 39g, Total Carbohydrates = 37g, Protein = 27g, Trans Fat = 1g, Iron Daily Value = 25%, Calcium Daily Value = 20%, dietary_fiber_g = 1g, Vitamin A Daily Value = 8% our model predicts:

```
##      fit      lwr      upr
## 1 618.5995 571.2563 665.9428
```

- The predicted calorie is **618.5995 calories**.
- The 95% prediction interval of calories content for the fast food with the given nutrients is between **571.2563** and **665.9428** calories.

4. Conclusion and Discussion

4.1. Approach

The approach is promising as we were able to create a multiple linear regression model that explained ~99% of the variability in Calories of menu items with a 23.9 calorie-difference between the fitted values and actual observed values.

Another variant of this approach can consider using even lesser predictors because from the stepwise selection process, the adjusted R-squared was as good as 97.5% using only the total fat and total carbohydrates predictors and reached 99% after adding the protein predictor.

4.2. Future Work

Follow-up work can be done to ensure that the selected model meets equal variance assumptions either through: - Box-Cox transformations: These work when the dependent variable values are all positive. We will need to explore menu items with 0 calories and determine whether they are significant to be kept in the data for the model. If they are not, we can remove them and perform Box-Cox transformations. If they are, then we can consider another approach such as - Weighted Least Squares Regression

References

1. Marcone, M. F., Madan, P., & Grodzinski, B. (2020). An overview of the sociological and environmental factors influencing eating food behavior in Canada. *Frontiers in Nutrition*, 7, 77. <https://doi.org/10.3389/fnut.2020.00077>
2. Polksy, J. Y., & Garriguet, D. (2021). Eating away from home in Canada: Impact on dietary intake. *Health Reports*, 32(8), 26-34. <https://doi.org/10.25318/82-003-x202100800003-eng>
3. Fast Food Nutrition. (n.d.). Canada - Fast food nutrition facts. Fast Food Nutrition. Retrieved from <https://fastfoodnutrition.org/ca/>
4. Arvidsson, J. (2023). Fast food nutrition: Nutritional values, micronutrients, and calories from six fast food restaurants [Dataset]. Kaggle. <https://www.kaggle.com/datasets/joebeachcapital/fast-food>
5. Lipman, D. (2023). MRModelPart1-Part4: Multiple linear regression [Lecture notes]. In Data603: Statistical Modelling with Data. University of Calgary. Based on content created by T. Ngamkham. Retrieved from <https://d2l.ucalgary.ca/>
6. Statsig. (2024, July 3). Understanding significance levels: A key to accurate data analysis. Retrieved from <https://www.statsig.com/perspectives/understanding-significance-levels-a-key-to-accurate-data-analysis>

Appendix

```
library(mosaic)
library(ggformula)
library(readr)
library(ggplot2)
library(gridExtra)
library(corrplot)
library(janitor)
library(mctest)
library(olsrr)
library(GGally)
library(lmtest)
library(knitr)
library(leaps)

options(scipen = 999)
set.seed(2021)

theme_set(theme_bw()) # change theme for ggplot2
trellis.par.set(theme=theme.mosaic()) # change default color scheme for lattice

knitr::opts_chunk$set(
  tidy=FALSE,
  size="small")

fast_food_df = read.csv('FastFoodNutritionScraped20241011.csv')
#head(fast_food_df)

#kable(head(fast_food_df[, 1:6]), format = "latex", longtable = TRUE)

fast_food_df = fast_food_df %>% janitor::clean_names() %>% filter(calories >= calories_from_fat)
#cat('The columns in the data include:\n')
#colnames(fast_food_df)
calories_violin_plot = ggplot(fast_food_df,
                               aes(x = company, y = calories, fill=company)) +
  geom_violin(alpha = 0.7) +
  geom_boxplot(width = 0.1, fill = "white", color = "black") +
  labs(title = "Calorie Content by Restaurant",
       x = NULL, y = "Calories", fill = "Company")
# calculate average calories per category
avg_calories = fast_food_df %>% group_by(category) %>%
  summarise(average_calories = mean(calories, na.rm = TRUE) )%>%
  arrange(average_calories)

# Convert Category to a factor to preserve sorted order in the plot
avg_calories$category <- factor(avg_calories$category, levels = avg_calories$category)

# visualization to see average calories per category
avg_category_calories = ggplot(avg_calories, aes(y = category, x = average_calories)) +
  geom_bar(stat = "identity", fill='blue') +
  labs(title = "Average Calories per Category", x = "Average Calories", y=NULL)
# visualization to see how trans fat is different based on company
trans_fat_violin_plot = ggplot(fast_food_df,
```



```

## companydairy-queen      FALSE   FALSE
## companymcdonalds       FALSE   FALSE
## companysubway           FALSE   FALSE
## companytim-hortons     FALSE   FALSE
## total_fat_g             FALSE   FALSE
## trans_fat_g             FALSE   FALSE
## cholesterol_mg          FALSE   FALSE
## total_carbohydrates_g  FALSE   FALSE
## dietary_fiber_g         FALSE   FALSE
## protein_g                FALSE   FALSE
## vitamin_a_daily_value   FALSE   FALSE
## vitamin_c_daily_value   FALSE   FALSE
## calcium_daily_value    FALSE   FALSE
## iron_daily_value        FALSE   FALSE
## 1 subsets of each size up to 11
## Selection Algorithm: exhaustive
##          companydairy-queen companymcdonalds companysubway companytim-hortons
## 1  ( 1 )   " "           " "           " "           " "
## 2  ( 1 )   " "           " "           " "           " "
## 3  ( 1 )   " "           " "           " "           " "
## 4  ( 1 )   "*"          " "           " "           " "
## 5  ( 1 )   "*"          " "           " "           " "
## 6  ( 1 )   "*"          " "           " "           " "
## 7  ( 1 )   "*"          " "           " "           " "
## 8  ( 1 )   "*"          " "           " "           " "
## 9  ( 1 )   "*"          " "           " "           " "
## 10 ( 1 )   "*"          "*"          " "           " "
## 11 ( 1 )   "*"          "*"          " "           " "
##          total_fat_g trans_fat_g cholesterol_mg total_carbohydrates_g
## 1  ( 1 )   "*"          " "           " "           " "
## 2  ( 1 )   "*"          " "           " "           "*"
## 3  ( 1 )   "*"          " "           " "           "*"
## 4  ( 1 )   "*"          " "           " "           "*"
## 5  ( 1 )   "*"          "*"          " "           "*"
## 6  ( 1 )   "*"          "*"          " "           "*"
## 7  ( 1 )   "*"          "*"          " "           "*"
## 8  ( 1 )   "*"          "*"          " "           "*"
## 9  ( 1 )   "*"          "*"          " "           "*"
## 10 ( 1 )   "*"          "*"          " "           "*"
## 11 ( 1 )   "*"          "*"          " "           "*"
##          dietary_fiber_g protein_g vitamin_a_daily_value vitamin_c_daily_value
## 1  ( 1 )   " "           " "           " "           " "
## 2  ( 1 )   " "           " "           " "           " "
## 3  ( 1 )   " "           "*"          " "           " "
## 4  ( 1 )   " "           "*"          " "           " "
## 5  ( 1 )   " "           "*"          " "           " "
## 6  ( 1 )   " "           "*"          "*"          " "
## 7  ( 1 )   " "           "*"          "*"          " "
## 8  ( 1 )   "*"          "*"          " "           " "
## 9  ( 1 )   "*"          "*"          "*"          " "
## 10 ( 1 )   "*"          "*"          "*"          " "
## 11 ( 1 )   "*"          "*"          "*"          " "
##          calcium_daily_value iron_daily_value
## 1  ( 1 )   " "           " "

```

```

## 2  ( 1 )   " "
## 3  ( 1 )   " "
## 4  ( 1 )   " "
## 5  ( 1 )   " "
## 6  ( 1 )   " "
## 7  ( 1 )   "*"
## 8  ( 1 )   "*"
## 9  ( 1 )   "*"
## 10 ( 1 )   "*"
## 11 ( 1 )   "*"

BIC=c(subsetModelSum$bic)
#cbind(BIC)

tTestModel = lm(calories ~ company + total_fat_g + trans_fat_g + cholesterol_mg +
                 total_carbohydrates_g + dietary_fiber_g + protein_g +
                 vitamin_a_daily_value + vitamin_c_daily_value + calcium_daily_value +
                 iron_daily_value,
                 data = mlr_df)
#summary(tTestModel)
caloriesAdditiveModel = lm(calories ~ total_fat_g + total_carbohydrates_g + protein_g +
                            trans_fat_g + iron_daily_value + calcium_daily_value +
                            dietary_fiber_g + company + vitamin_a_daily_value,
                            data = mlr_df)
#summary(caloriesAdditiveModel)
interacModel = lm(calories ~ (total_fat_g + total_carbohydrates_g + protein_g +
                               trans_fat_g + iron_daily_value + calcium_daily_value +
                               dietary_fiber_g + company + vitamin_a_daily_value)^2,
                  data = mlr_df)
#summary(interacModel)
interacModelv2 = lm(calories ~ total_fat_g + total_carbohydrates_g + protein_g +
                     trans_fat_g + iron_daily_value + calcium_daily_value +
                     dietary_fiber_g + company + vitamin_a_daily_value +
                     total_fat_g:total_carbohydrates_g + total_fat_g:calcium_daily_value +
                     total_fat_g:company + total_fat_g:vitamin_a_daily_value +
                     total_carbohydrates_g:trans_fat_g +
                     total_carbohydrates_g:iron_daily_value +
                     total_carbohydrates_g:calcium_daily_value +
                     total_carbohydrates_g:vitamin_a_daily_value +
                     protein_g:iron_daily_value + protein_g:calcium_daily_value +
                     trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
                     iron_daily_value:calcium_daily_value +
                     iron_daily_value:dietary_fiber_g + iron_daily_value:company +
                     iron_daily_value:vitamin_a_daily_value +
                     calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
                     dietary_fiber_g:company,
                     data = mlr_df)
#summary(interacModelv2)
caloriesInteracModel = lm(calories ~ total_fat_g + total_carbohydrates_g + protein_g +
                           trans_fat_g + iron_daily_value + calcium_daily_value +
                           dietary_fiber_g + company + vitamin_a_daily_value +
                           total_fat_g:total_carbohydrates_g + total_fat_g:calcium_daily_value +
                           total_fat_g:company + total_fat_g:vitamin_a_daily_value +
                           total_carbohydrates_g:trans_fat_g +

```

```

total_carbohydrates_g:iron_daily_value +
total_carbohydrates_g:calcium_daily_value +
total_carbohydrates_g:vitamin_a_daily_value +
protein_g:calcium_daily_value +
trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
iron_daily_value:dietary_fiber_g + iron_daily_value:company +
calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
dietary_fiber_g:company,
      data = mlr_df)
#summary(caloriesInteracModel)
yCorrDf = mlr_df %>% select(calories, total_fat_g, total_carbohydrates_g, protein_g,
                               trans_fat_g, iron_daily_value, calcium_daily_value,
                               dietary_fiber_g, company, vitamin_a_daily_value)
#ggpairs(yCorrDf, lower = list(continuous = "smooth_loess",
#higherOrderModel = lm(calories ~ total_fat_g + I(total_fat_g^2) + total_carbohydrates_g + protein_g +
#                      trans_fat_g + iron_daily_value + calcium_daily_value +
#                      dietary_fiber_g + company + vitamin_a_daily_value +
#                      total_fat_g:total_carbohydrates_g + total_fat_g:calcium_daily_value +
#                      total_fat_g:company + total_fat_g:vitamin_a_daily_value +
#                      total_carbohydrates_g:trans_fat_g +
#                      total_carbohydrates_g:iron_daily_value +
#                      total_carbohydrates_g:calcium_daily_value +
#                      total_carbohydrates_g:vitamin_a_daily_value +
#                      protein_g:calcium_daily_value +
#                      trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
#                      iron_daily_value:dietary_fiber_g + iron_daily_value:company +
#                      calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
#                      dietary_fiber_g:company,
#                      data = mlr_df)
#summary(higherOrderModel)
higherOrderModelv2 = lm(calories ~ total_fat_g + I(total_carbohydrates_g^2) + total_carbohydrates_g + p
                        trans_fat_g + iron_daily_value + calcium_daily_value +
                        dietary_fiber_g + company + vitamin_a_daily_value +
                        total_fat_g:total_carbohydrates_g + total_fat_g:calcium_daily_value +
                        total_fat_g:company + total_fat_g:vitamin_a_daily_value +
                        total_carbohydrates_g:trans_fat_g +
                        total_carbohydrates_g:iron_daily_value +
                        total_carbohydrates_g:calcium_daily_value +
                        total_carbohydrates_g:vitamin_a_daily_value +
                        protein_g:calcium_daily_value +
                        trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
                        iron_daily_value:dietary_fiber_g + iron_daily_value:company +
                        calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
                        dietary_fiber_g:company,
                        data = mlr_df)
#summary(higherOrderModelv2)
CaloriesHigherOrderModel = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) +
                             trans_fat_g + iron_daily_value + calcium_daily_value +
                             dietary_fiber_g + company + vitamin_a_daily_value +
                             total_fat_g:calcium_daily_value +
                             total_fat_g:vitamin_a_daily_value +
                             total_carbohydrates_g:trans_fat_g +
                             total_carbohydrates_g:iron_daily_value +

```

```

total_carbohydrates_g:calcium_daily_value +
total_carbohydrates_g:vitamin_a_daily_value +
protein_g:calcium_daily_value +
trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
iron_daily_value:dietary_fiber_g + iron_daily_value:company +
calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
dietary_fiber_g:company,
      data = mlr_df)
#summary(CaloriesHigherOrderModel)
higherOrderModelv4 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + I(
  trans_fat_g + iron_daily_value + calcium_daily_value +
  dietary_fiber_g + company + vitamin_a_daily_value +
  total_fat_g:calcium_daily_value +
  total_fat_g:vitamin_a_daily_value +
  total_carbohydrates_g:trans_fat_g +
  total_carbohydrates_g:iron_daily_value +
  total_carbohydrates_g:calcium_daily_value +
  total_carbohydrates_g:vitamin_a_daily_value +
  protein_g:calcium_daily_value +
  trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
  iron_daily_value:dietary_fiber_g + iron_daily_value:company +
  calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
  dietary_fiber_g:company,
      data = mlr_df)
#summary(higherOrderModelv4)
higherOrderModelv5 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + p(
  trans_fat_g + iron_daily_value + calcium_daily_value +
  dietary_fiber_g + company + vitamin_a_daily_value +
  total_fat_g:calcium_daily_value +
  total_fat_g:vitamin_a_daily_value +
  total_carbohydrates_g:trans_fat_g +
  total_carbohydrates_g:iron_daily_value +
  total_carbohydrates_g:calcium_daily_value +
  total_carbohydrates_g:vitamin_a_daily_value +
  protein_g:calcium_daily_value +
  trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
  iron_daily_value:dietary_fiber_g + iron_daily_value:company +
  calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
  dietary_fiber_g:company,
      data = mlr_df)
#summary(higherOrderModelv5)
higherOrderModelv6 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + p(
  trans_fat_g + iron_daily_value + calcium_daily_value + I(calcium_daily_value^2) +
  dietary_fiber_g + company + vitamin_a_daily_value +
  total_fat_g:calcium_daily_value +
  total_fat_g:vitamin_a_daily_value +
  total_carbohydrates_g:trans_fat_g +
  total_carbohydrates_g:iron_daily_value +
  total_carbohydrates_g:calcium_daily_value +
  total_carbohydrates_g:vitamin_a_daily_value +
  protein_g:calcium_daily_value +
  trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
  iron_daily_value:dietary_fiber_g + iron_daily_value:company +
  calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
  dietary_fiber_g:company +
  protein_g:calcium_daily_value +
  trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
  iron_daily_value:dietary_fiber_g + iron_daily_value:company +
  calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
  dietary_fiber_g:company,
      data = mlr_df)
#summary(higherOrderModelv6)

```

```

    calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
    dietary_fiber_g:company,
    data = mlr_df)
#summary(higherOrderModelv6)

higherOrderModelv7 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + p
trans_fat_g + I(trans_fat_g^2) + iron_daily_value + calcium_daily_value +
dietary_fiber_g + company + vitamin_a_daily_value +
total_fat_g:calcium_daily_value +
total_fat_g:vitamin_a_daily_value +
total_carbohydrates_g:trans_fat_g +
total_carbohydrates_g:iron_daily_value +
total_carbohydrates_g:calcium_daily_value +
total_carbohydrates_g:vitamin_a_daily_value +
protein_g:calcium_daily_value +
trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
iron_daily_value:dietary_fiber_g + iron_daily_value:company +
calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
dietary_fiber_g:company,
data = mlr_df)

#summary(higherOrderModelv7)

higherOrderModelv8 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + p
trans_fat_g + iron_daily_value + I(iron_daily_value^2) + calcium_daily_value +
dietary_fiber_g + company + vitamin_a_daily_value +
total_fat_g:calcium_daily_value +
total_fat_g:vitamin_a_daily_value +
total_carbohydrates_g:trans_fat_g +
total_carbohydrates_g:iron_daily_value +
total_carbohydrates_g:calcium_daily_value +
total_carbohydrates_g:vitamin_a_daily_value +
protein_g:calcium_daily_value +
trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
iron_daily_value:dietary_fiber_g + iron_daily_value:company +
calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
dietary_fiber_g:company,
data = mlr_df)

#summary(higherOrderModelv8)

higherOrderModelv9 = lm(calories ~ total_fat_g + total_carbohydrates_g + I(total_carbohydrates_g^2) + p
trans_fat_g + iron_daily_value + calcium_daily_value +
dietary_fiber_g + I(dietary_fiber_g^2) + company + vitamin_a_daily_value +
total_fat_g:calcium_daily_value +
total_fat_g:vitamin_a_daily_value +
total_carbohydrates_g:trans_fat_g +
total_carbohydrates_g:iron_daily_value +
total_carbohydrates_g:calcium_daily_value +
total_carbohydrates_g:vitamin_a_daily_value +
protein_g:calcium_daily_value +
trans_fat_g:iron_daily_value + trans_fat_g:vitamin_a_daily_value +
iron_daily_value:dietary_fiber_g + iron_daily_value:company +
calcium_daily_value:dietary_fiber_g + calcium_daily_value:company +
dietary_fiber_g:company,
data = mlr_df)

#summary(higherOrderModelv9)

```

```

#plot(caloriesAdditiveModel, which = 1)
#plot(caloriesAdditiveModel, which = 3)
#bptest(caloriesAdditiveModel)
#plot(caloriesAdditiveModel, which = 2)
#shapiro.test(residuals(caloriesAdditiveModel))
#plot(caloriesAdditiveModel, which=5)

#confint(caloriesAdditiveModel, level = 0.95)

#range(mlr_df$total_fat_g)
#range(mlr_df$total_carbohydrates_g)
#range(mlr_df$protein_g)
#range(mlr_df$trans_fat_g)
#range(mlr_df$iron_daily_value)
#range(mlr_df$calcium_daily_value)
#range(mlr_df$dietary_fiber_g)
#range(mlr_df$company)
#range(mlr_df$vitamin_a_daily_value)

#mlr_df[32, ]

# Actual Calorie = 620
newData = data.frame(total_fat_g = 39, total_carbohydrates_g = 37, protein_g = 27, trans_fat_g = 1, iron_g = 1, calcium_g = 300, dietary_fiber_g = 10, company = "Company A", vitamin_a_daily_value = 500)
#predict(caloriesAdditiveModel,newdata = newData,interval = "predict")

```