Nutritional Facts for McDonald’s menu

Runtime Terror

Joshua Phillips   
*CSE*  
*PES University, EC campus*Bangalore, India  
joshua.phillips899@gmail.com

Ketan Malempati  
*CSE*  
*PES University, EC campus* Bangalore, India  
ketanmalempati8h@gmail.com

S Deepashree  
*CSE*  
*PES University, EC campus*Bangalore, India  
deepasrees2000@gmail.com

***Abstract*— The paper we have chosen for our project is the Nutritional Facts for McDonald’s menu. In our review paper we wanted to analyze the menu in the McDonald’s menu all around the world with respect to the calories, cholesterol level, etc with respect to each item in the menu.**

***Keywords—Nutrition, Regression***

# INTRODUCTION

The Fast food Industry has changed the lives of everything drastically in the last 40 years or so. Arguably one of the first fast-food restaurants White Castle opened in 1921 but it was only when Ray Croc stepped in and wanted to change and invest in something he thought would revolutionise the industry forever. The McDonald brothers opened their first restaurant adjacent to the Monrovia Airport in 1937. It was a tiny octagonal building informally called The Airdrome. That octagonal building was later moved to 1398 North E Street in San Bernardino, California in 1940. It was originally a barbecue drive-in, but the brothers discovered that most of their profits came from hamburgers. In 1948, they closed their restaurant for three months, reopening it in December as a walk-up hamburger stand that sold hamburgers, potato chips, and orange juice; the following year, french fries and Coca-Cola were added to the menu. This simplified menu, and food preparation using assembly line principles, allowed them to sell hamburgers for 15 cents, or about half as much as at a sit-down restaurant. What was just a single restaurant that was started by the McDonald’s brothers, he wanted to invest in their ‘speedee system’ and open up another franchise in Des Plaines, Illinois in April, 1955.

Kroc purchased the company outright in 1961, and his strict operational guidelines helped transform McDonald's into the world’s largest restaurant franchise before his death in 1984, at the age of 81.Ray Kroc’s vision was that there would be 1,000 McDonald’s restaurants solely in the United States. Yet, McDonald’s continued to grow and expand into international markets beginning in 1967 opening in Canada and Puerto Rico. Today, the company has over 36,000 restaurants in over 100 nations.

# LITERATURE SURVEY

## **Paper 1**

1. Paper Details

*Title :*Literature Review of Correlation matrix

*Authors:* Michael Friendly

1. Summary

a) *Abstract:*

This paper describes a set of techniques we subsume under the name “corrgram”, based on the rendering the value of a correlation to depict its sign and magnitude. A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses.

b) *Claims:*

The correlation matrix plays an important role in multivariate analysis since by itself it captures the pairwise degrees of relationship between different components of a random vector. We can claim to have presented a more general and comprehensive account of the possibilities than

has appeared previously. We have also (a) suggested a new scheme for ordering variables in such displays, (b) extended the idea of correlation mapping to more general concepts of dependence and independence, and (c) illustrated (we hope convincingly) why they might be useful.

## **Paper 2**

1. Paper Details

*Title:* Review of Regression Analysis Models

*Authors:* Dr. Amita Goel, Aviral Gupta, and Akshay Sharma

*Year:* 2017

1. Summary

*a) Abstract*:

Regression Analysis is a technique to find out the relationship between different variables. Regression looks

closely into how a dependent variable is affected upon varying an independent variable while keeping the

other independent variables constant. Linear Regression Analysis model is used to fit linear data, a Polynomial Regression Analysis model focuses on a data set representing polynomial relationship between data parameters.

b) *Claims:*

Regression is a really effective tool in statistical analysis of data. Different Regression models are used in different situations. A Linear Regression Model is best suited to data representing a linear relationship between two variables, a Polynomial Regression Model is used in case of multiple variables having a polynomial relationship.

***C. Paper 3***

1. Paper Details

*Title :* Data Analysis on ‘Nutrition Facts for McDonald's Menu’ Data-set using Python

*Authors:*Neha Tiwari , Prof. Vaishali Gatty

*Year:*2017

1. Summary

*a) Abstract:*

It uses open source platforms and libraries such as NumPy, Scipy, matplotlib, pandas, scikit learn etc. This paper aims to highlight data analysis of ' Nutrition Facts for McDonald's Menu' dataset using Python. The Indian food industry has risen as a high-development and high-benefit area because of its huge potential for esteem expansion, especially inside the food processing industry. This dataset is used to analyze nutritious and non-nutritious food items in the menu. It uses various python libraries to analyze this dataset to represent the data in the form of different charts.

*b) Claims:*

It is analyzed that the items in the menu dataset can be categorized as nutritious food and non-nutritious food based on different chart diagrams and range values in percentage. It is beneficial for demonstrating different range values for food nutrients such as vitamin A, vitamin B, vitamin C, sugar, dietary fibers, fats, carbohydrates, cholesterol, iron, sodium and protein for their proper consumption from menu items.

# PRELIMINARY ANALYSIS

## Outliers and Missing values

The dataset that we had chosen did not have any missing values as such, therefore to perform cleaning on the dataset we ourselves removed a few values and worked on it to get a clean dataset.

## Units

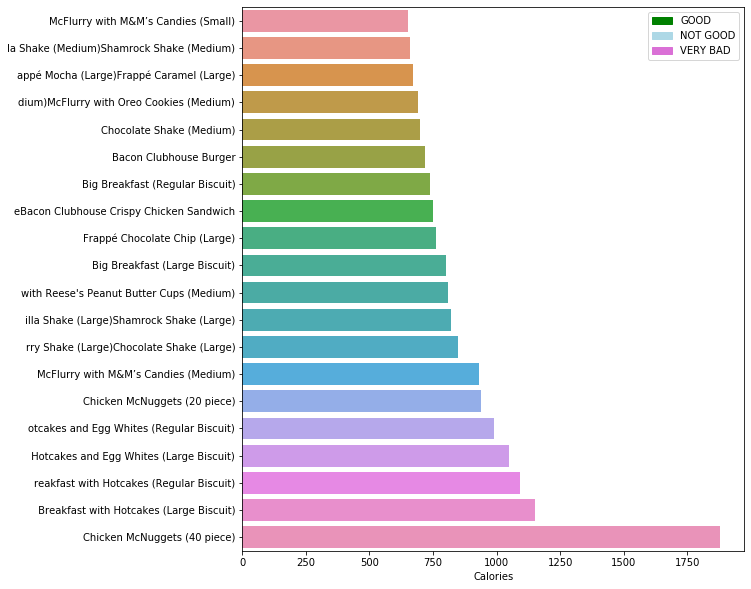
The units that are being used in the dataset are calories

## Normalisation

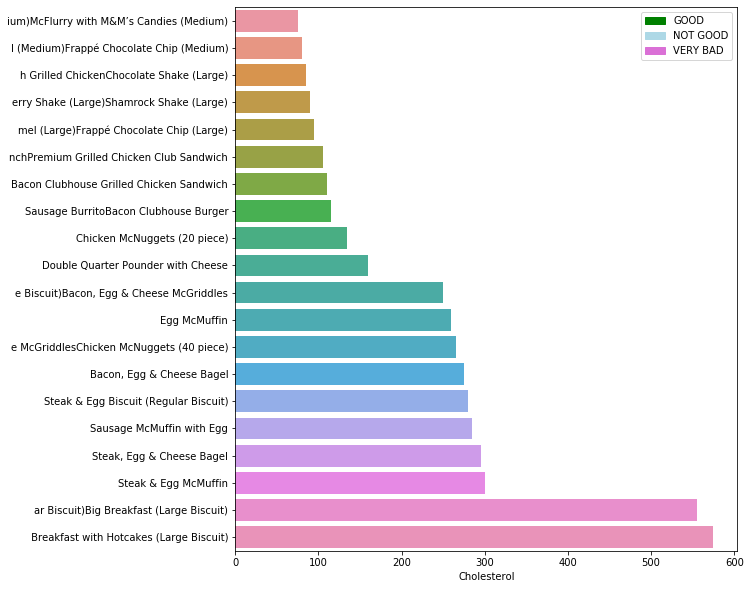
* In our dataset we have not normalised any of the values yet with respect to the the unit that they are in because we haven’t compared any of the data yet and though there are outliers we feel that they are necessary in our set because we need to know as in the case like the highest amount of calories is by 40 piece Chicken McNugget whilst the lowest being Diet Coke (small).

# DATASET INSPECTION

## Bar Graph

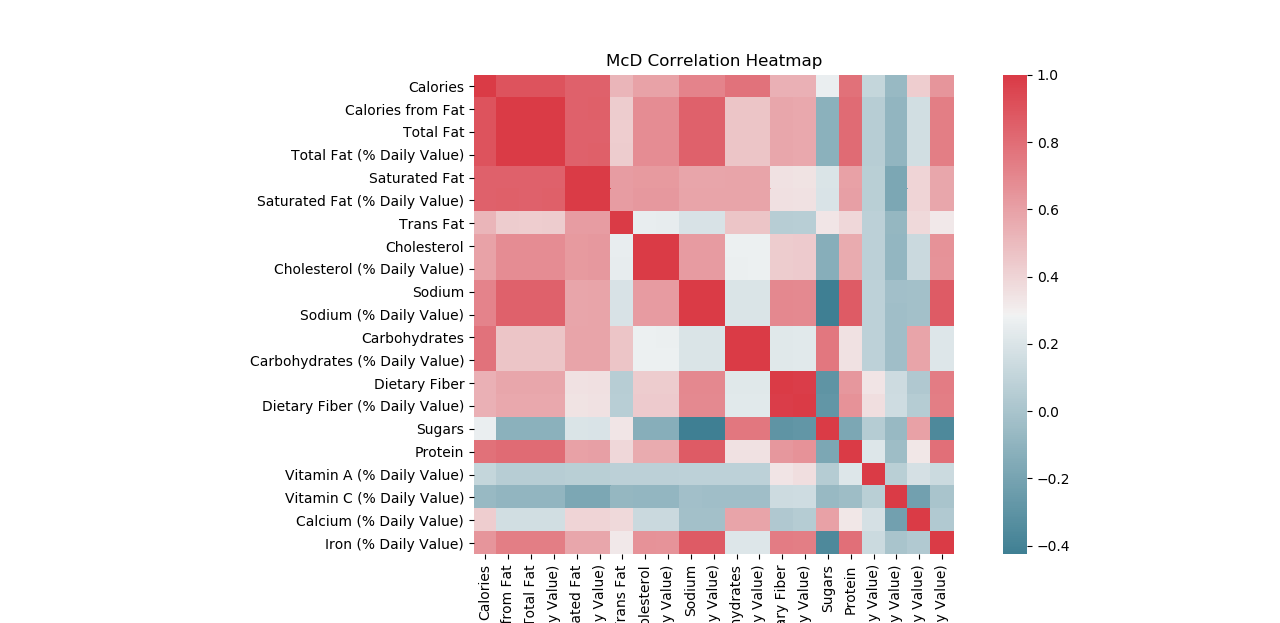


/\*mostly only deserts and shakes have the highest calories\*/

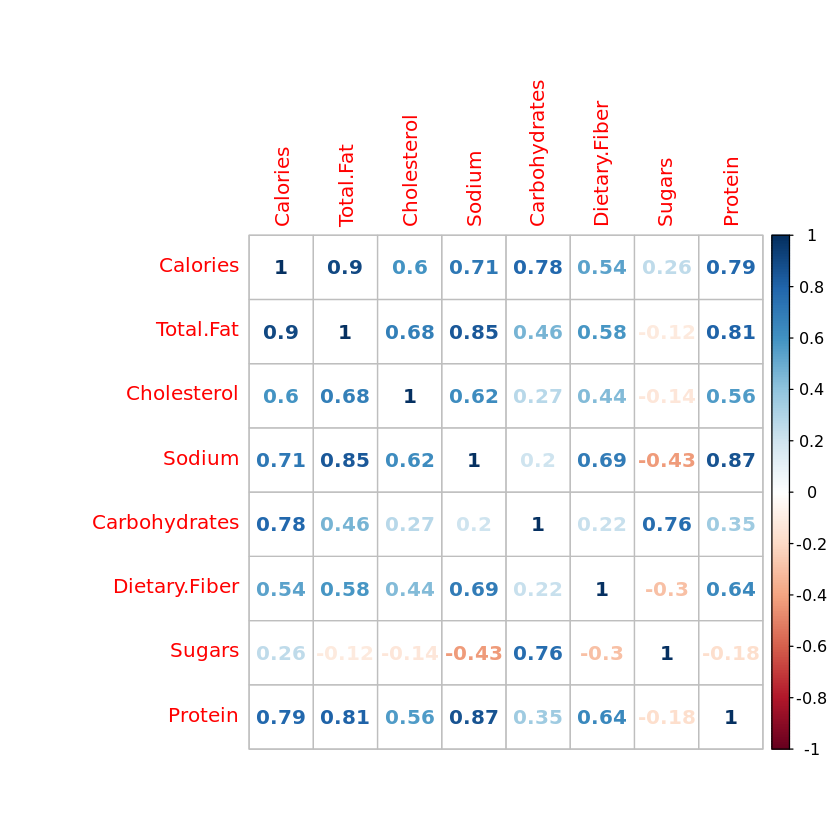


From this above visualisation, we can note that the highest amount of calories are in deserts and shakes. While the sugar itself cannot be blamed for having high-calorie content another plausible fact is that with high amounts of sugar content are consumed, a) people are now going to consume food that has high-calorie content. b) people may be more likely to consume more of the food since it tastes a little better. When people consume food containing added sugar, they tend to consume more calories. However, when people consume sugar just by itself, they do not consume large amounts of it. This is explained in the following visualisation and that is why sugar has a negative correlation with calories when it’s expected to be a high positive correlation with calories.

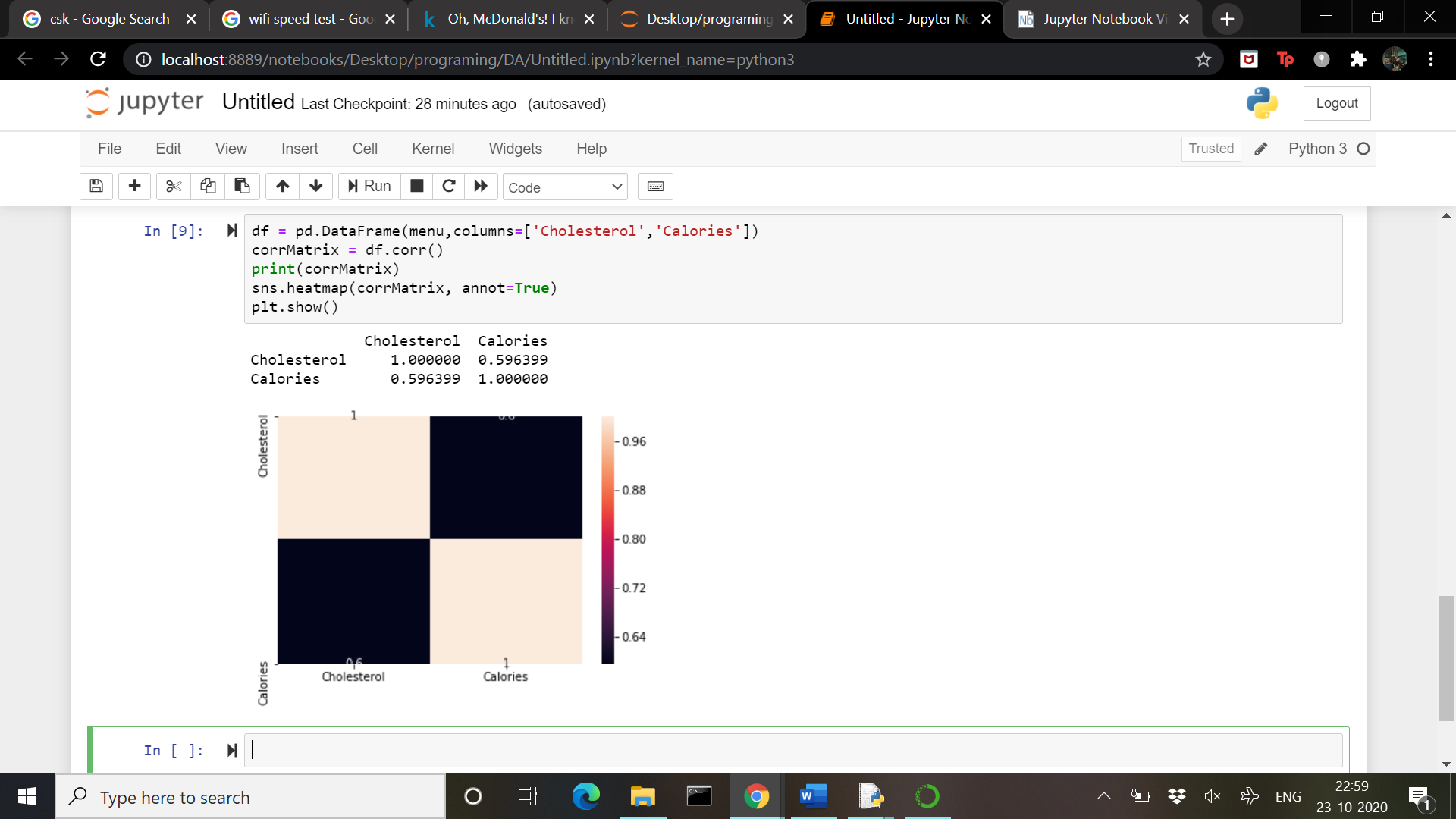
## Heatmaps



This above image is just a heatmap showing how strong the correlations are between all the values that we have in our dataset. Other than the fact that obviously they have strong correlations with themselves we can notice that values such as Iron and Sodium, Protein and Sodium have strong correlations.

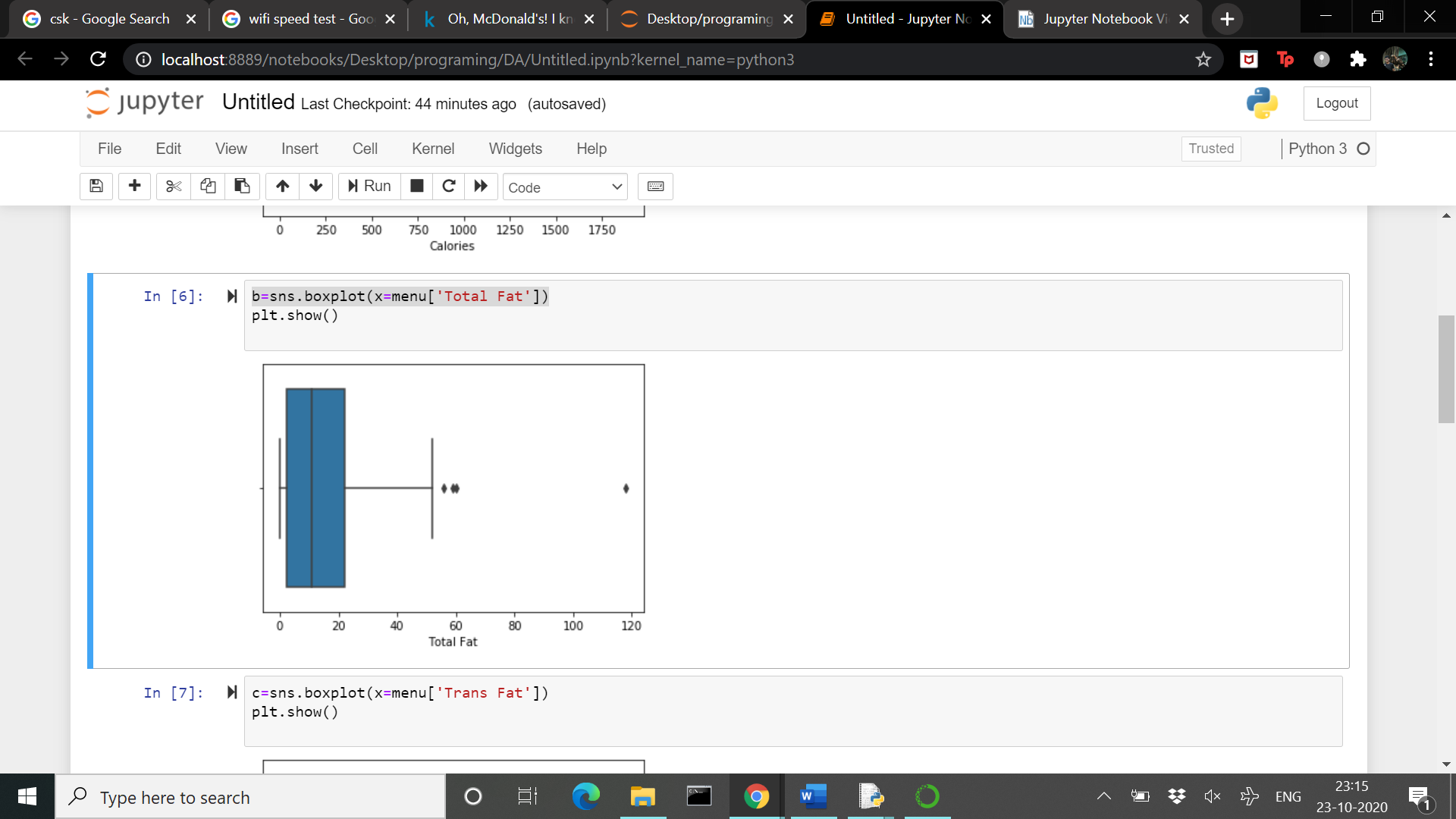
C.  *Correlation Matrix*

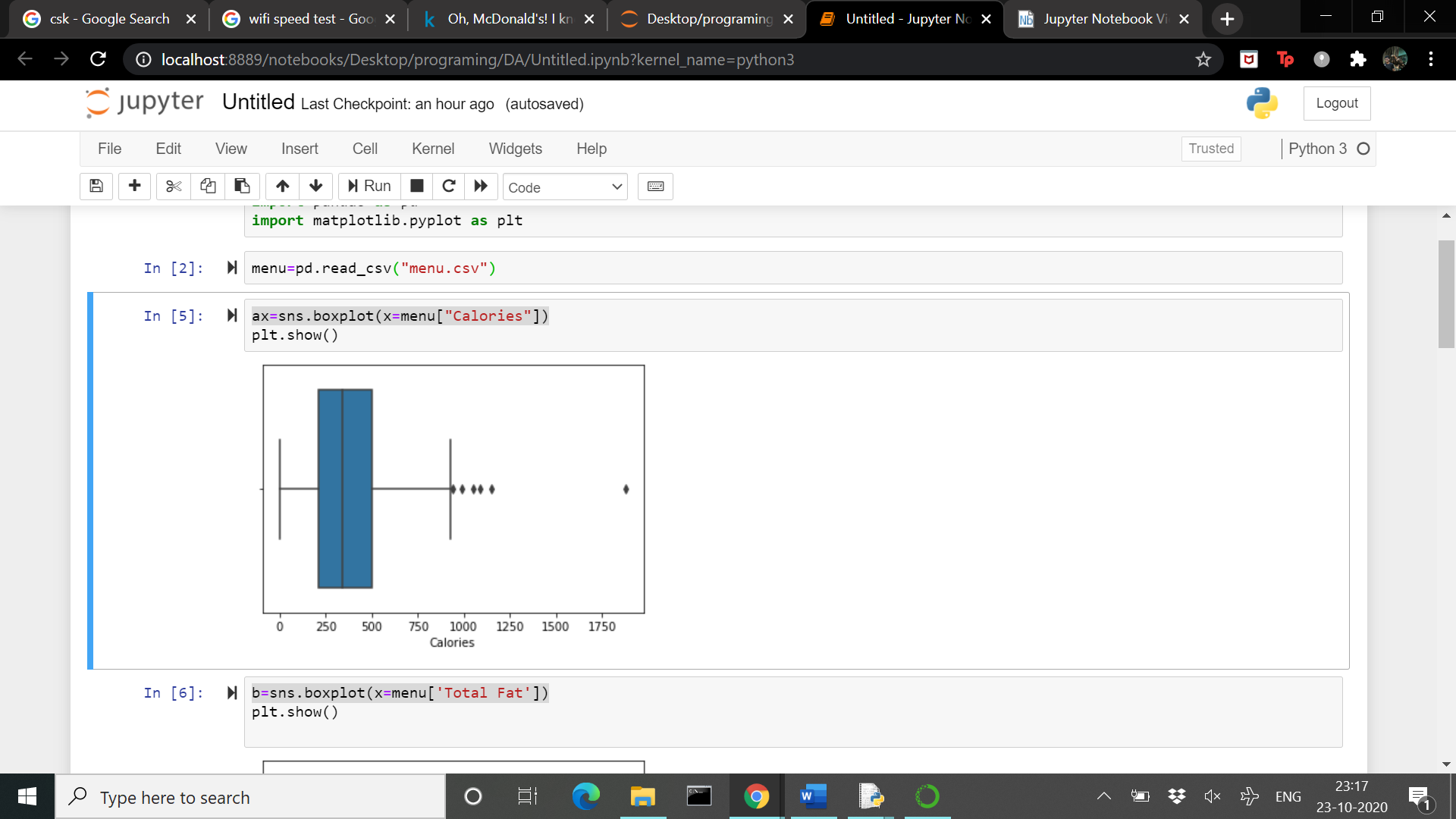
Total fat has a high correlation = 0.9, followed by protein = 0.79 and carbohydrates = 0.78

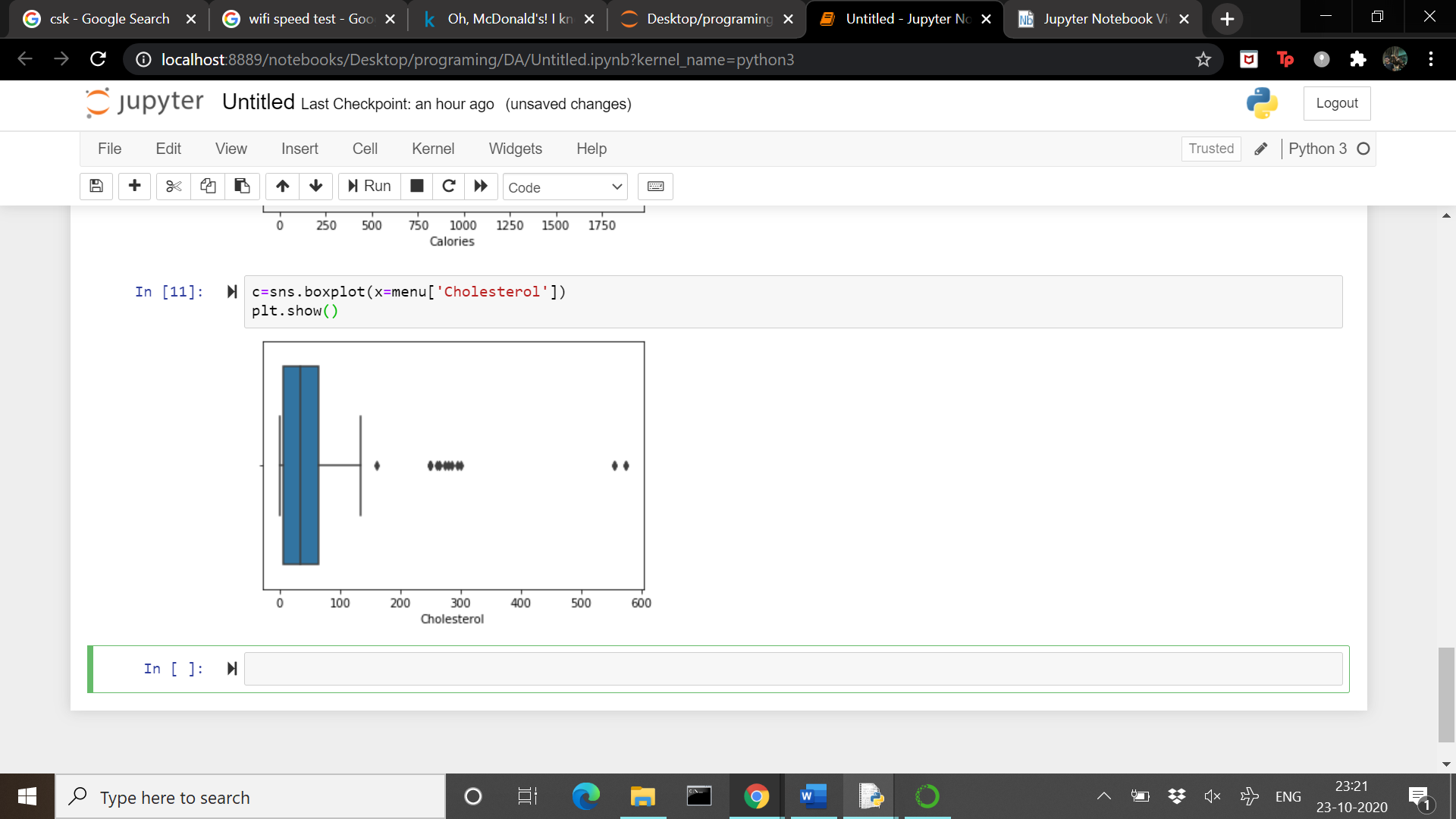


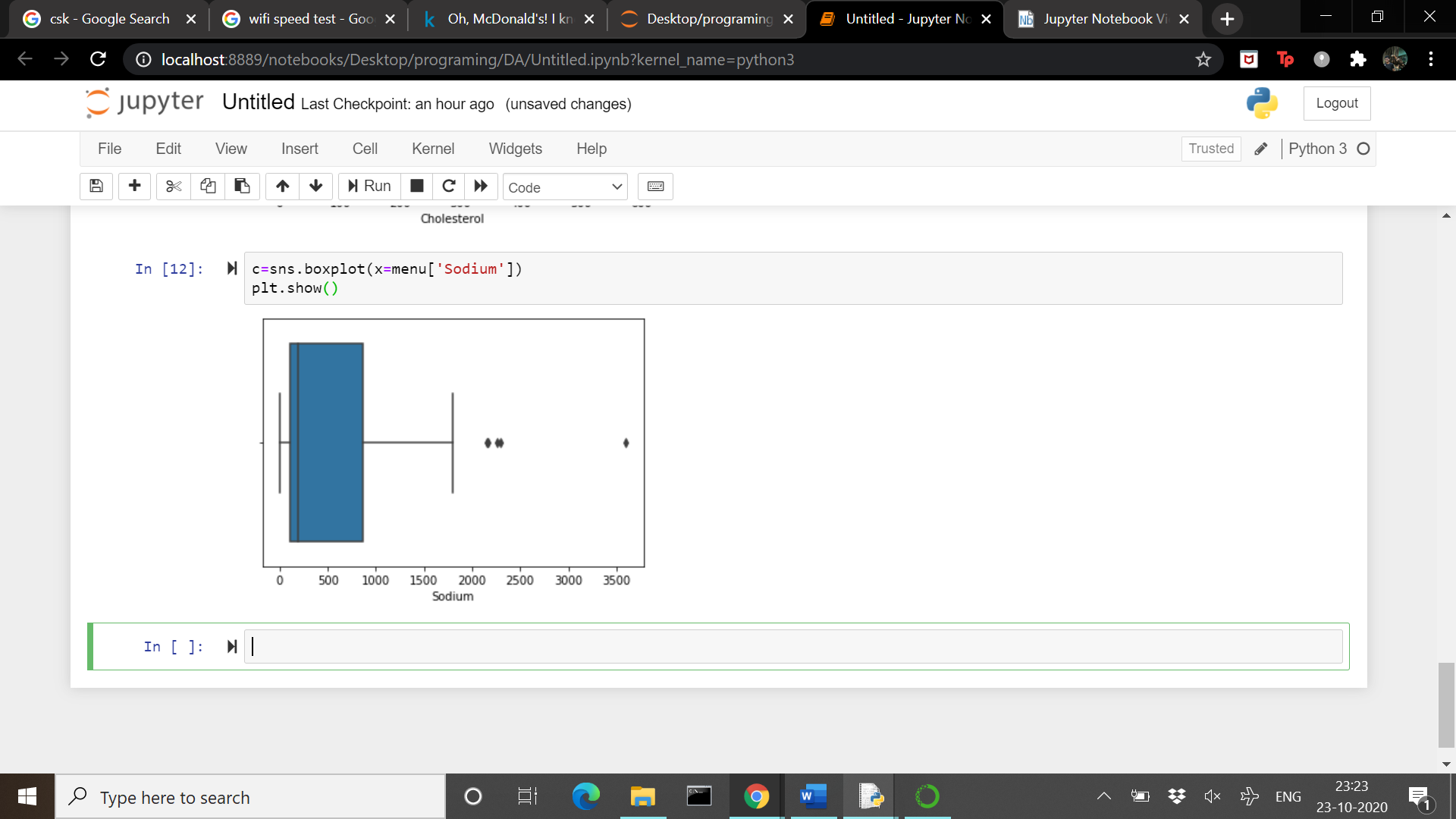
More specifically we have plotted the heatmap between Cholesterol vs Calories.

## D. Box Plots

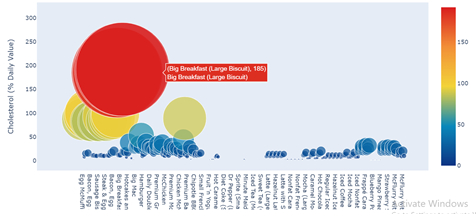






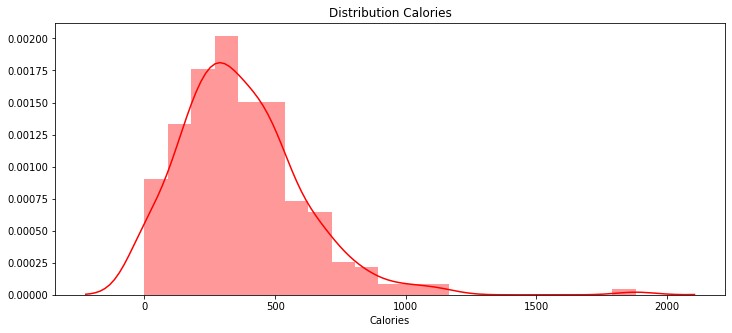
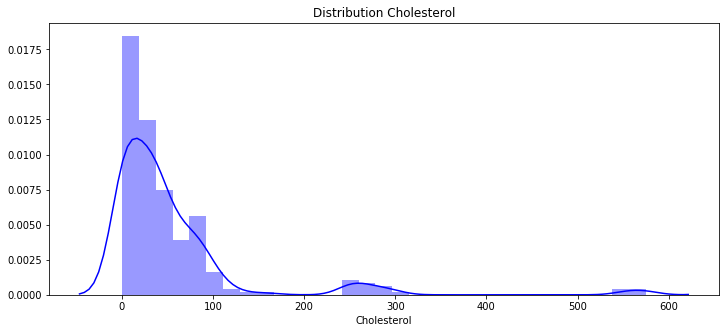


*E. Bubble Plot*



The most striking visuals from the scatter plot are the few large red plots close to the top. These circular plots were scaled such that the higher the Cholesterol (% Daily value), the larger the plot thereby making for intuitive visuals. As we can see, the main culprit (red circle) is the MacDonald's Big Breakfast range, accounting for a whooping 185% of Cholesterol (% Daily value).

Further down from larger red plots, there are a greater number of yellow circular plots which can be attributed to items such as the Egg/Sausage McMuffin range contributing to nearly a day's worth of Cholesterol.

*F. Histogram*

V. PREDICTION MODEL

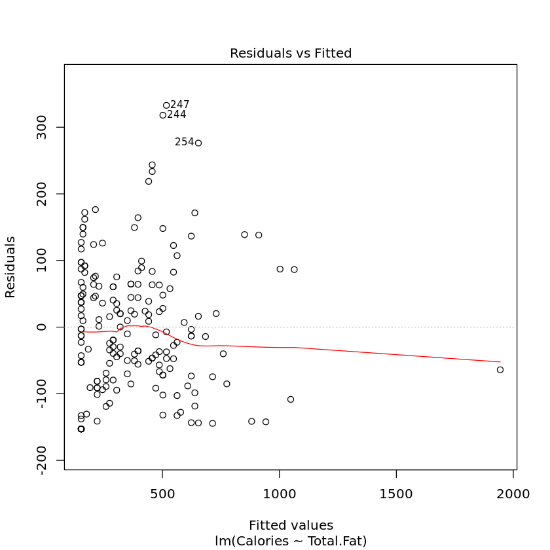
In our project we are going to explore two different models.

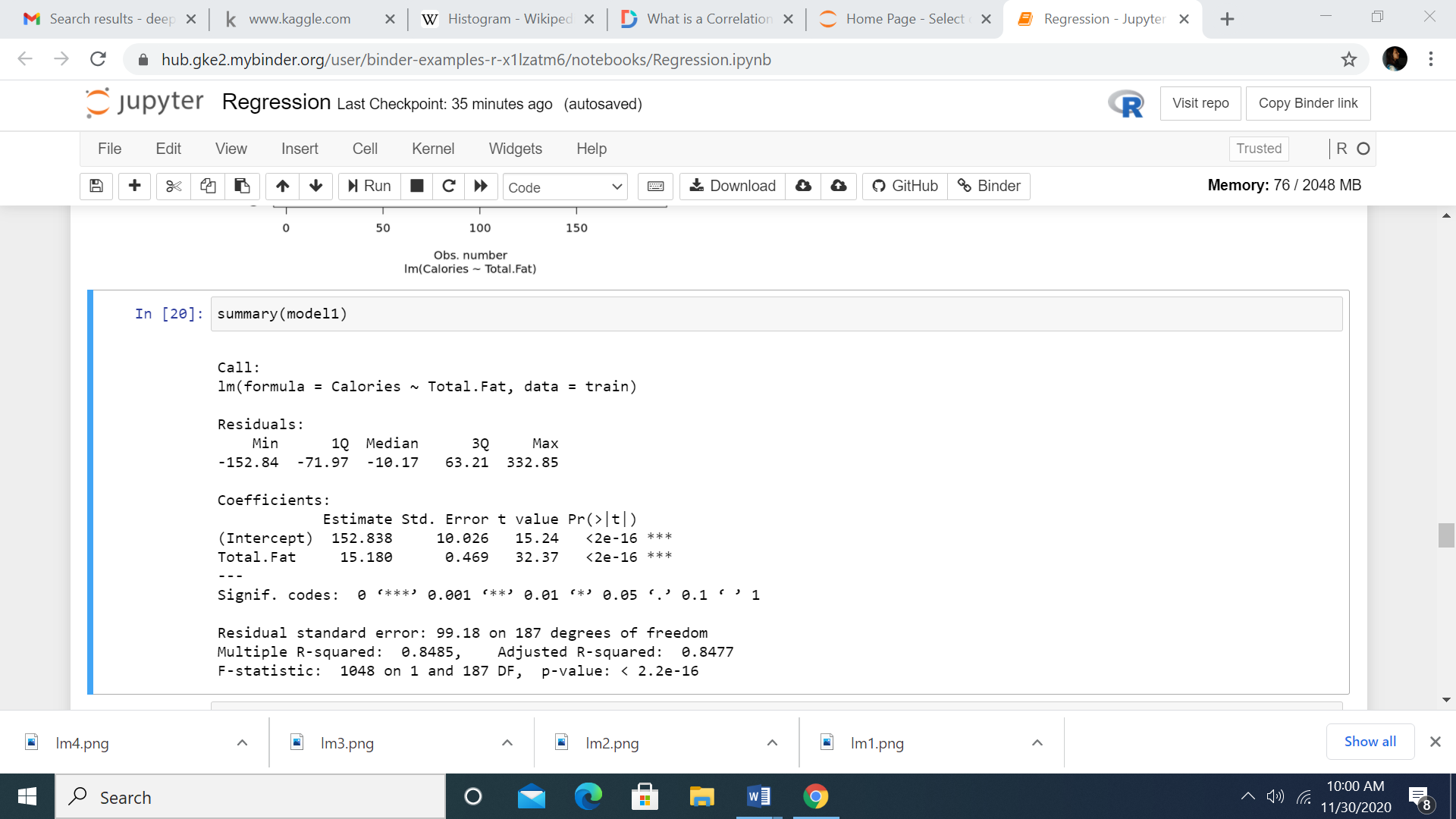
1. Regression Models
2. Random Forest Method
3. REGRESSION MODEL

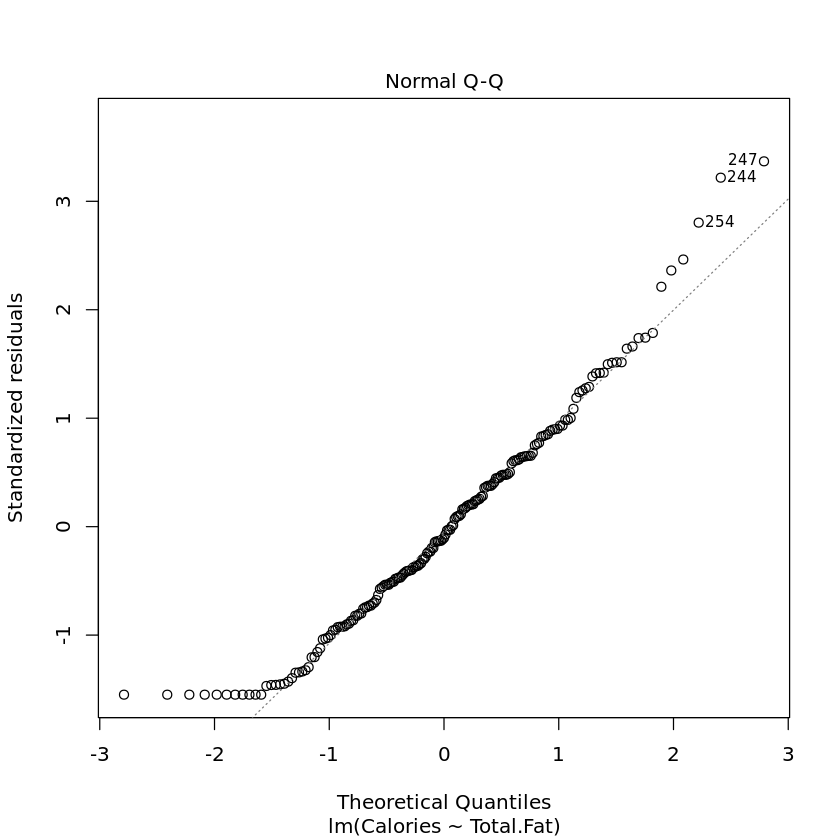
***Linear Regression Model*** between calories and total fat.

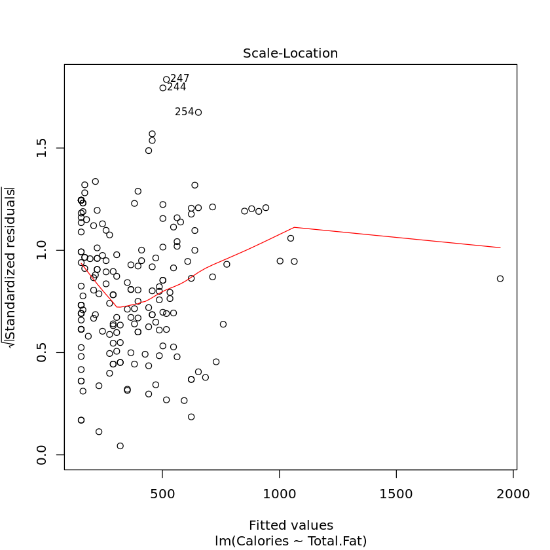
Dependent variable -- Calories

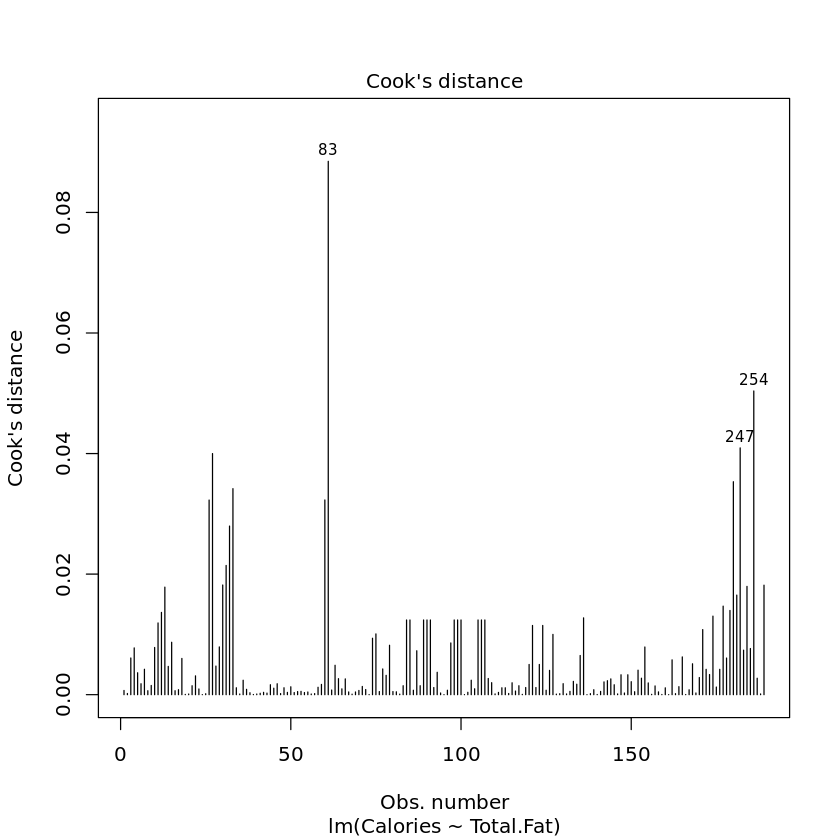
Independent variable -- Total fat











Value of intercept = 152.838 and value of slope = 15.180

Both the values are significant R-squared = 85%

The overall p-value is also significant

The linear equation to predict calories:

*Calories = 152.838 + 15.180\*total fat*

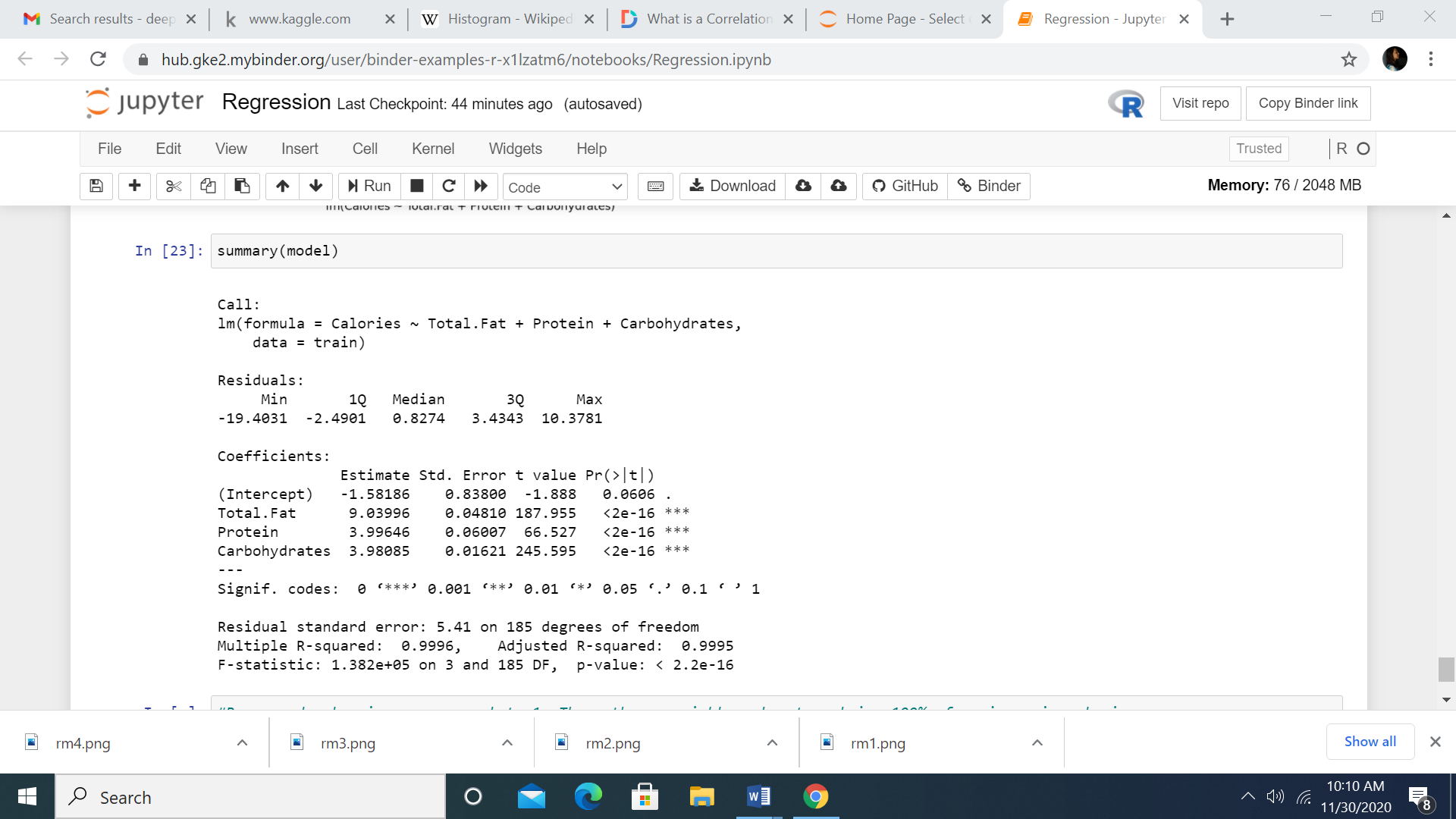
***Multiple Regression Model***

(From corrplot we found out that total fat, protein and carbohydrates are highly correlated.)

Dependent variable -- Calories

Independent Variables -- Total Fat + Protein + Carbohydrates

##### 



##### 

##### 

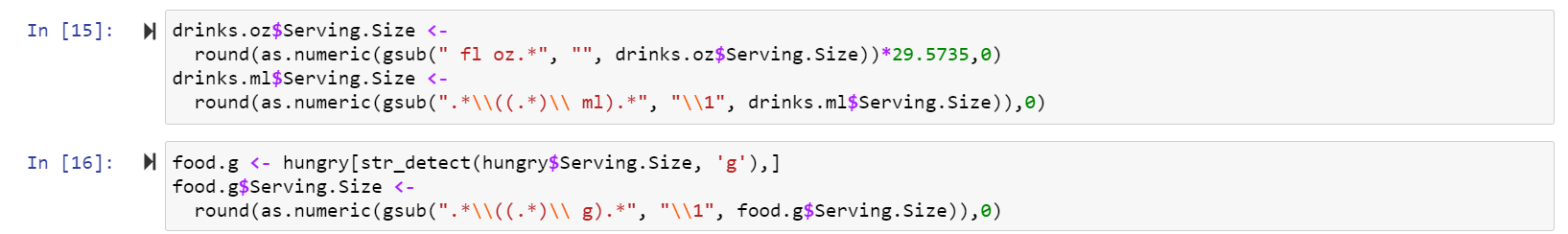
##### 

R-squared value is approx equal to 1. These three variables almost explain 100% of variance in calories.

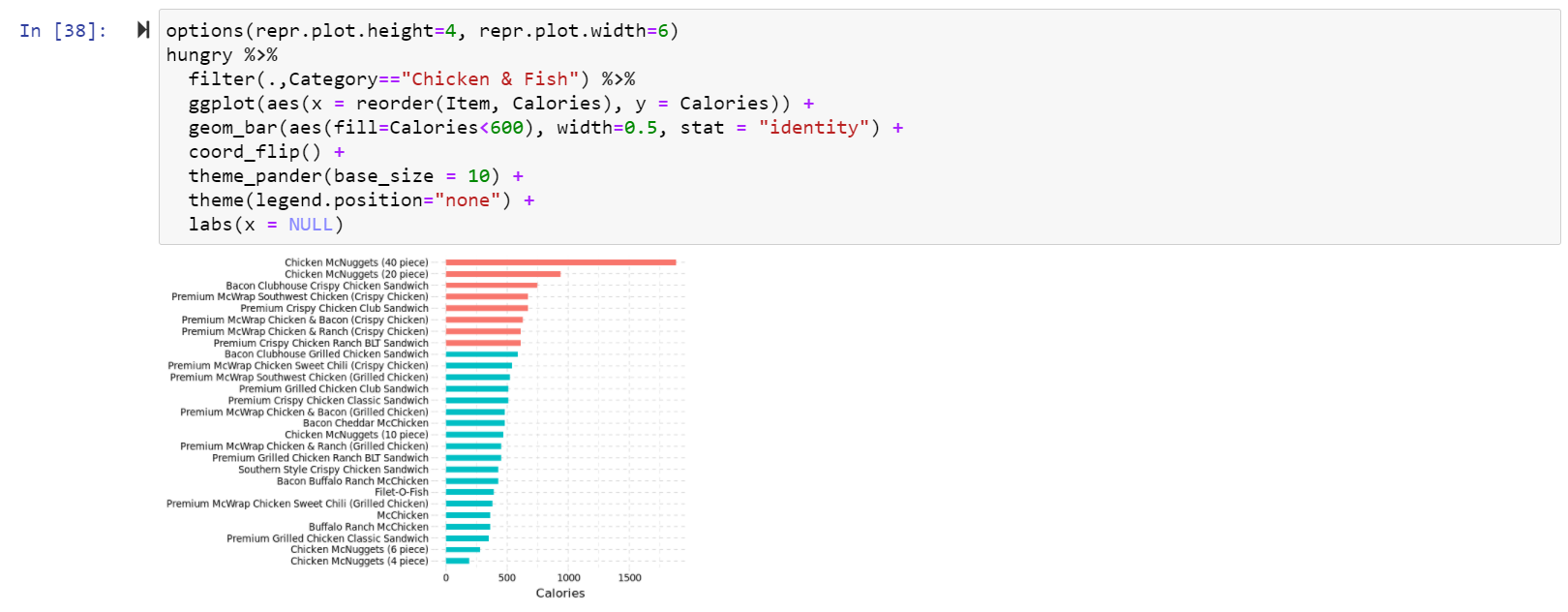
1. Random Forest Method

Using this particular method we are going to find out what are the important variables to predict how many calories an item has.

We start by converting all the units of drinks to ml, and all foods to grams.



Next we plotted to see what items have the highest amounts of calories

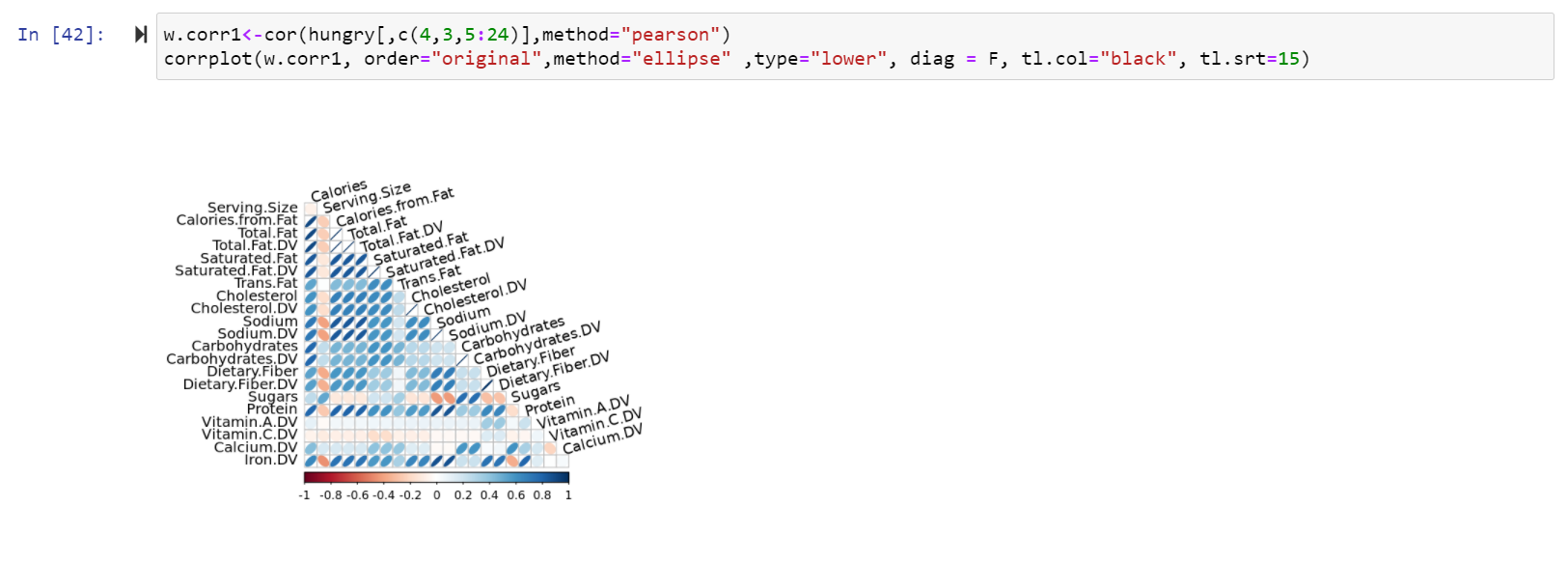


The point to note is that chicken nuggets come in the highest and lowest in the amount of calories that it has, but the difference between the high and low is the number of pieces in a meal. So there is some relation with the size of each item with the amount of calories it has as well.

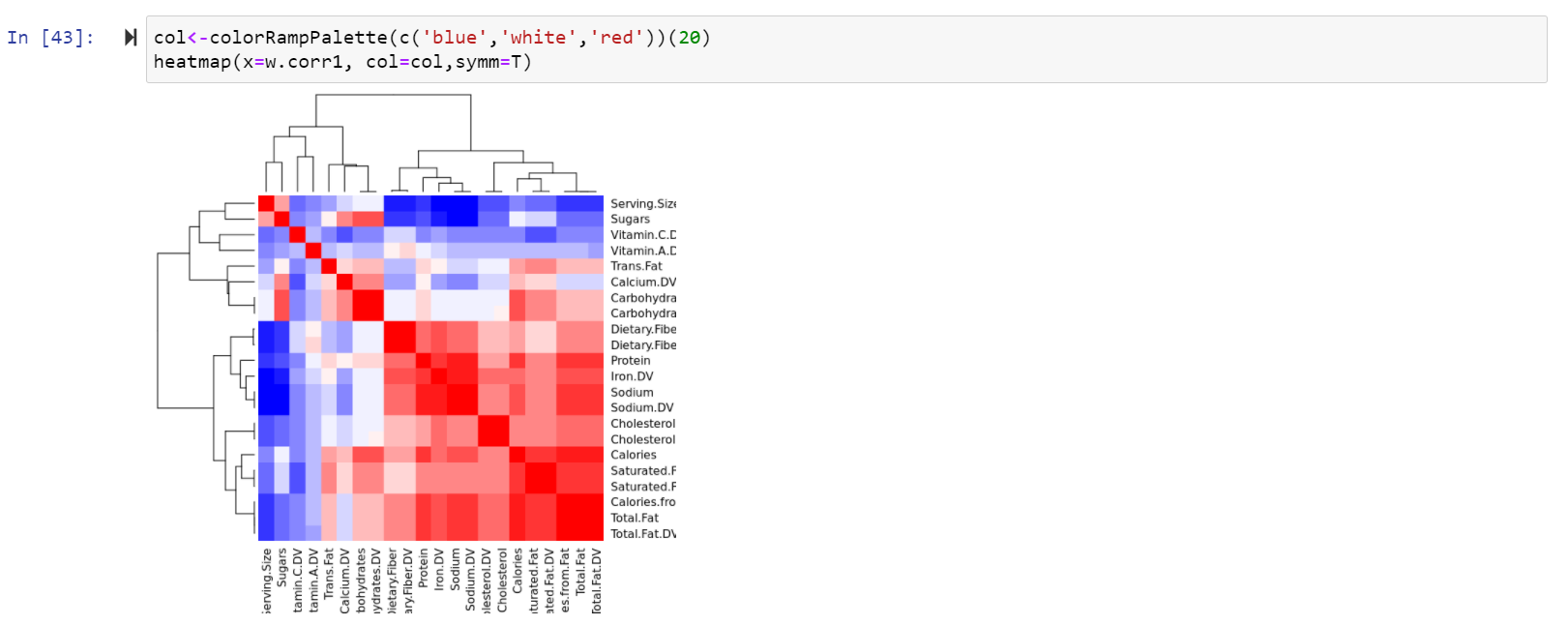
The above point is justified in this graph below as well



Next we find a correlation plot to check what columns affect the calorie amount.



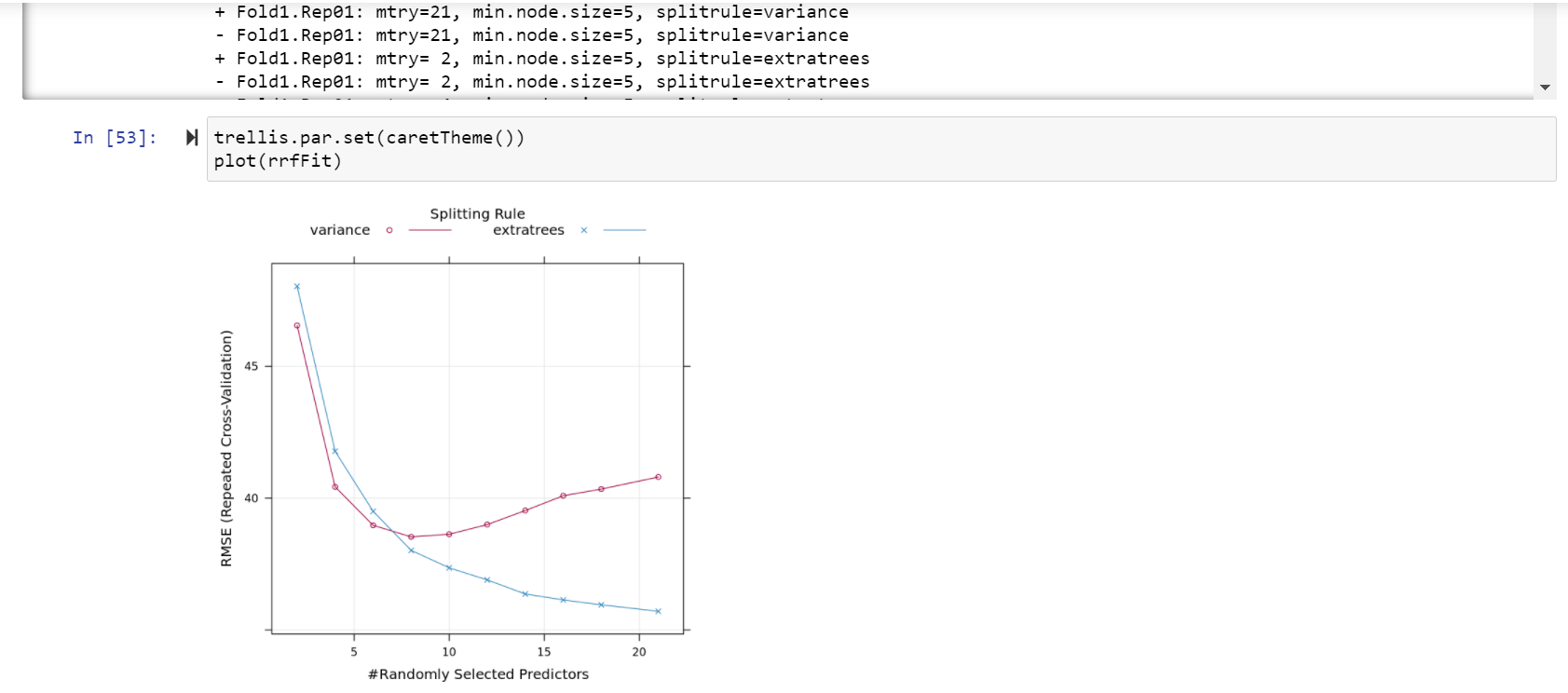
We then plot a heatmap to find out what are related



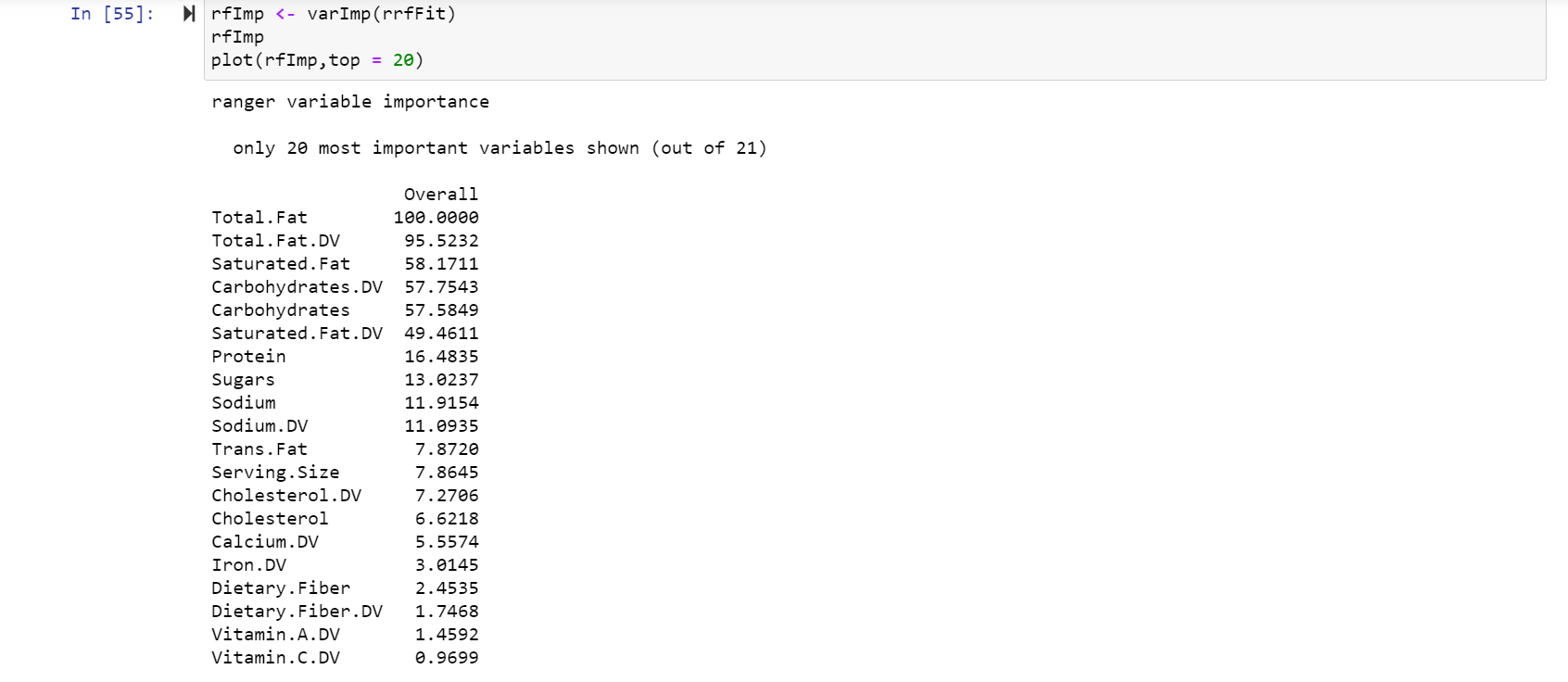
Now we train the dataset with 70% for training and the rest for testing.

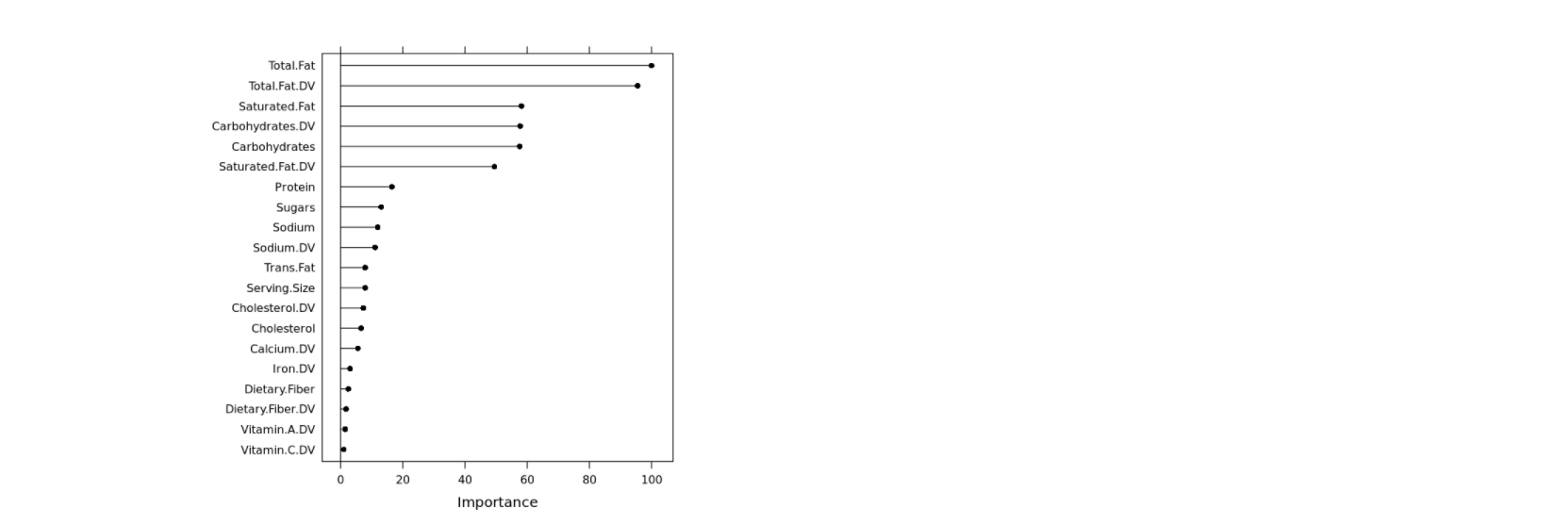
Next we start off with our Random Forest.

We train our Random Forest initially using the ranger method in ‘caret’ and after testing we plot it.



After our Analysis we find out the most important variables that contribute to the amount of calories that an item has.





In this we can see that Total Fat, Saturated Fat, Carbohydrates are the most influential variables in the amount of calories that an item has. It does not mean that the rest of the variables don’t have any effect but they are almost negligible when compared to these three.

VI. CONCLUSION

Most fast-food joints are always unhealthy with high content of salt, sugar etc. Due to this and the fact that it boomed, especially in America, obesity has been a major issue due to the number of people eating in fast-food joints. There are many data-sets and values through which we can find out how important each variable in food it and how it directly or indirectly affects how many calories that we eat but in our project we have specifically chosen the McDonald’s Menu and using this as a reference found out what items we need to cut down in order to reduce our calorie intake.

Therefore, if you want to be a health freak and control the amount of calorie intake that you have for every single meal, most important variables when predicting the calories are Total fat, Saturated fat and carbohydrates. So, if your menu is full of total fat, saturated fat and carbohydrates, you may exceed your daily calories.