McDonalds Case Study

-by Aruneema

About the case study

In this case study we are provided with a dataset that has the details about the food served in McDonald's menu. The objective it to do basic exploratory analysis and find basic insights into the nutritional content of the food served and find that:

1. **Which food categories have the highest and lowest varieties**
2. **Which all variables have an outlier**
3. **Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?**
4. **Which item contributes maximum to the Sodium intake?**
5. **Which 4 food items contain the most amount of Saturated Fat?**

ABOUT THE DATA

We are presented with the dataset containing information about different category of food items served at McDonalds and their nutrient contents. The data has following variable:

- Category- there are 9 categories of food i.e. Breakfast, Beef & Pork, Chicken & Fish, Salads, Snacks & Slides, Desserts, Beverages, Coffee & Tea and Smoothies and Shake.
- Item- Each category of food has many items under it. There are total 260 items.
- Serving Size
- Calories
- Calories from Fat
- Total Fat
- Total Fat (% Daily Value)
- Saturated Fat
- Saturated Fat (% Daily Value)
- Trans Fat
- Cholesterol
- Cholesterol (% Daily Value)
- Sodium
- Sodium (% Daily Value)
- Carbohydrates
- Carbohydrates (% Daily Value)
- Dietary Fiber
- Dietary Fiber (% Daily Value)
- Sugars
- Protein
- Vitamin A (% Daily Value)
- Vitamin C (% Daily Value)
- Calcium (% Daily Value)
- Iron (% Daily Value)

EXPLORATORY DATA ANALYSIS

The data set for the given case looks as follows:

	Category	Item	Serving Size	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	...	Carbohydrates	Carbohydrates (% Daily Value)	Dietary Fiber	Dietary Fiber (% Daily Value)	Sugars
0	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13.0	20	5.0	25	0.0	...	31	10	4	17	3
1	Breakfast	Egg White Delight	4.8 oz (135 g)	250	70	8.0	12	3.0	15	0.0	...	30	10	4	17	3
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200	23.0	35	8.0	42	0.0	...	29	10	4	17	2
3	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)	450	250	28.0	43	10.0	52	0.0	...	30	10	4	17	2
4	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)	400	210	23.0	35	8.0	42	0.0	...	30	10	4	17	2

5 rows × 24 columns

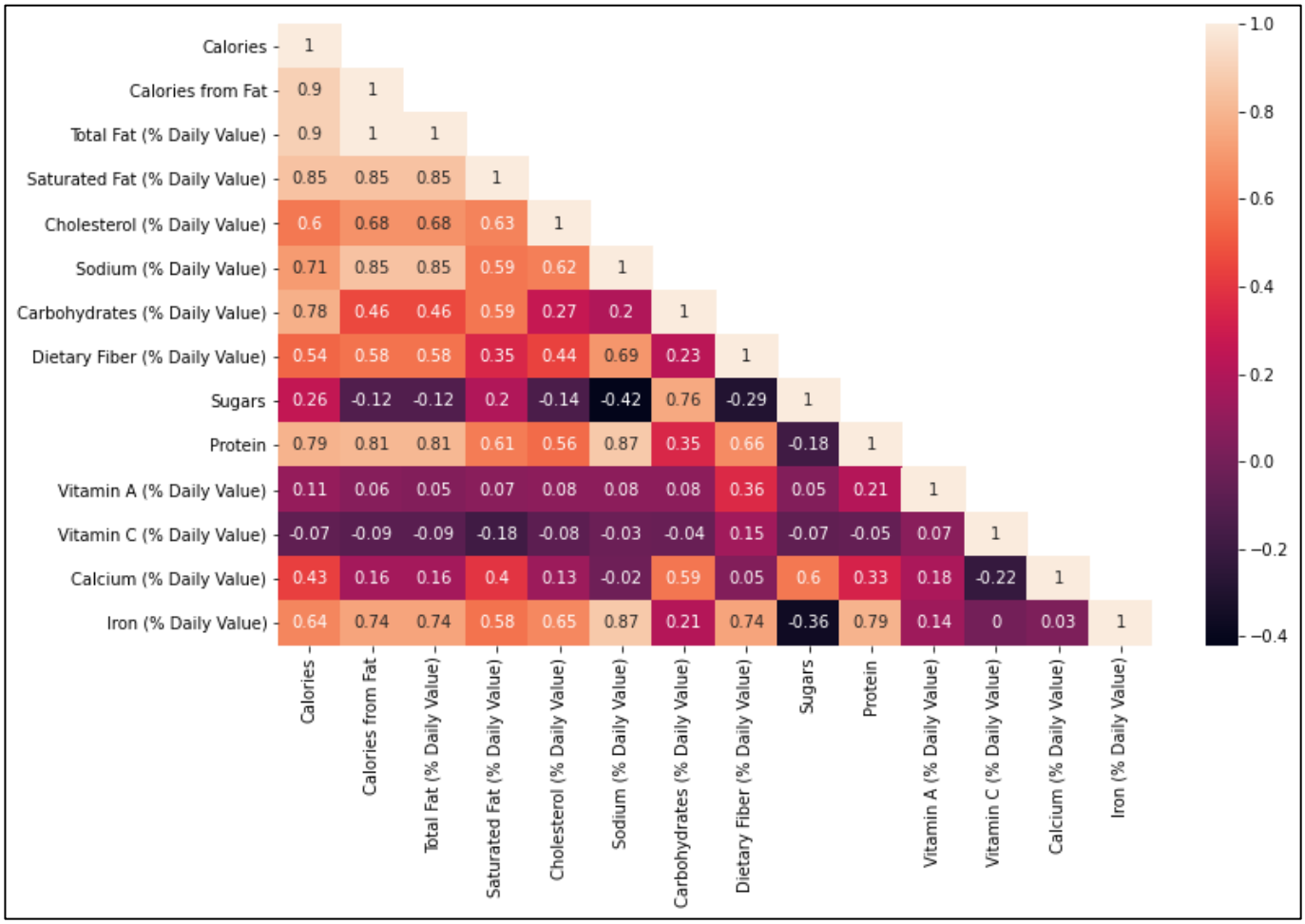
Table 1: Dataframe of the given dataset

	count	mean	std	min	25%	50%	75%	max
Calories	260.0	368.27	240.27	0.0	210.00	340.0	500.00	1880.0
Calories from Fat	260.0	127.10	127.88	0.0	20.00	100.0	200.00	1060.0
Total Fat	260.0	14.17	14.21	0.0	2.38	11.0	22.25	118.0
Total Fat (% Daily Value)	260.0	21.82	21.89	0.0	3.75	17.0	35.00	182.0
Saturated Fat	260.0	6.01	5.32	0.0	1.00	5.0	10.00	20.0
Saturated Fat (% Daily Value)	260.0	29.97	26.64	0.0	4.75	24.0	48.00	102.0
Trans Fat	260.0	0.20	0.43	0.0	0.00	0.0	0.00	2.5
Cholesterol	260.0	54.94	87.27	0.0	5.00	35.0	65.00	575.0
Cholesterol (% Daily Value)	260.0	18.39	29.09	0.0	2.00	11.0	21.25	192.0
Sodium	260.0	495.75	577.03	0.0	107.50	190.0	865.00	3600.0
Sodium (% Daily Value)	260.0	20.68	24.03	0.0	4.75	8.0	36.25	150.0
Carbohydrates	260.0	47.35	28.25	0.0	30.00	44.0	60.00	141.0
Carbohydrates (% Daily Value)	260.0	15.78	9.42	0.0	10.00	15.0	20.00	47.0
Dietary Fiber	260.0	1.63	1.57	0.0	0.00	1.0	3.00	7.0
Dietary Fiber (% Daily Value)	260.0	6.53	6.31	0.0	0.00	5.0	10.00	28.0
Sugars	260.0	29.42	28.68	0.0	5.75	17.5	48.00	128.0
Protein	260.0	13.34	11.43	0.0	4.00	12.0	19.00	87.0
Vitamin A (% Daily Value)	260.0	13.43	24.37	0.0	2.00	8.0	15.00	170.0
Vitamin C (% Daily Value)	260.0	8.53	26.35	0.0	0.00	0.0	4.00	240.0
Calcium (% Daily Value)	260.0	20.97	17.02	0.0	6.00	20.0	30.00	70.0
Iron (% Daily Value)	260.0	7.73	8.72	0.0	0.00	4.0	15.00	40.0

Table 2: Data Description

Vitamin C (% Daily Value)	5.502231
Vitamin A (% Daily Value)	4.580331
Cholesterol (% Daily Value)	3.804200
Total Fat (% Daily Value)	2.162011
Calories from Fat	2.145508
Protein	1.579924
Sodium (% Daily Value)	1.545986
Calories	1.452498
Iron (% Daily Value)	1.187908
Dietary Fiber (% Daily Value)	1.154245
Sugars	1.031940
Carbohydrates (% Daily Value)	0.903599
Saturated Fat (% Daily Value)	0.685296
Calcium (% Daily Value)	0.593555
dtype: float64	

Table 3: Skewness of Numeric variables



Plot 1: Correlation heatmap

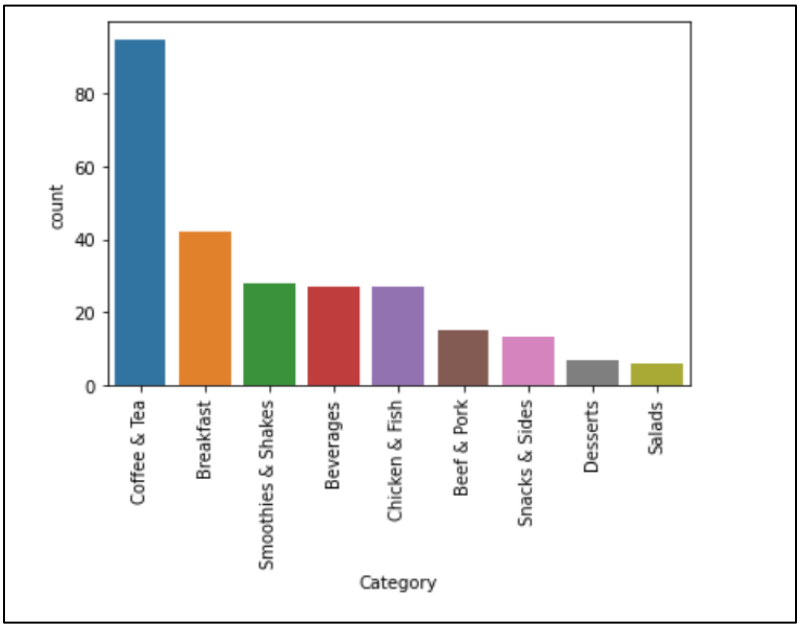
After uploading the data in Jupyter, basic EDA steps were performed to understand the shape, structure, type and statistical information about the data. It was found that-

A) There were no null values in the data
B) It consisted of 260 rows and 24 columns with the following descriptive statistics
C) Most of the variables are integer type, except for 'Category', 'Item' & 'Serving Size' which are objects and 'Total fat', 'Saturated Fat' and 'Trans Fat' being float.

D) The standard deviation of calories and sodium is quite high
E) All the numeric variables seem to be rightly skewed
F) Plot 1 suggest strong correlation between-
    - Saturated fats and 'Calories from fat' and 'Total Fat'
    - Sodium and 'Calories from fat' and 'Total Fat'
    - Protein and 'Calories from fat', 'Total Fat' and 'Sodium'

Q1. Graphically which food categories have the highest and lowest varieties.

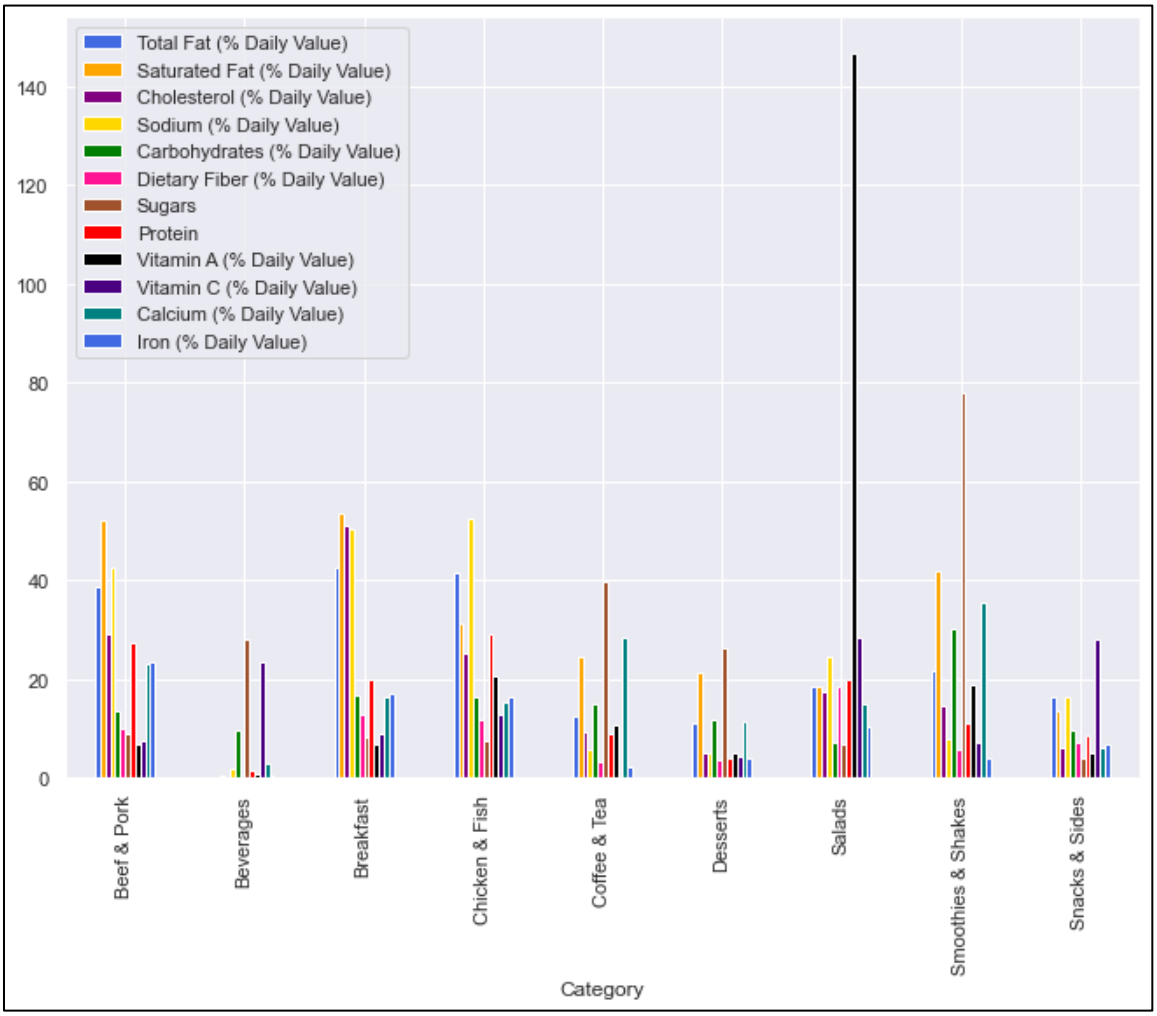
Upon plotting a count plot on variable 'Categories', Coffee and Tea has highest variety and salads have lowest variety.



Plot 2: Number of varieties for each Category

The food group (carbohydrates, proteins, fats, vitamins and minerals) composition of each category in plot 3, which tells us that-

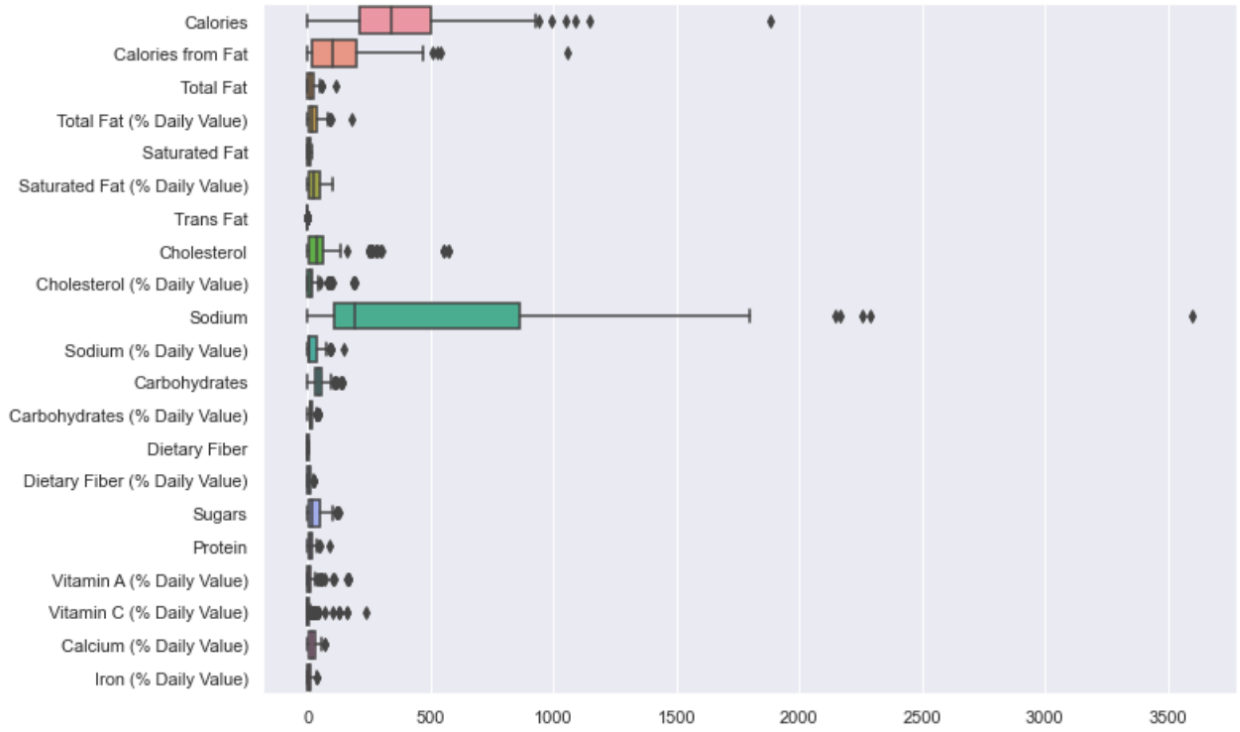
- Most of the food category are characterised by high average value of sugar, sodium and saturated fats content.
- Salads seem to have very high vitamin A mean value, while Smoothies and Shakes have relatively high sugar content.
- Beef & Pork, Breakfast, Chicken & Fish and Salads, relative have high average vitamins levels



Plot 3: Food group content in each category

Q2. Which all variables have an outlier?

On plotting boxplot of each variable, it seems that almost all the variables have outliers. However, the variable 'Sodium' has the most spread out outlier, with 'Chicken McNuggets' being the outer most outlier. Even in 'Calories' Chicken McNuggets is the outer most (right most) outlier.



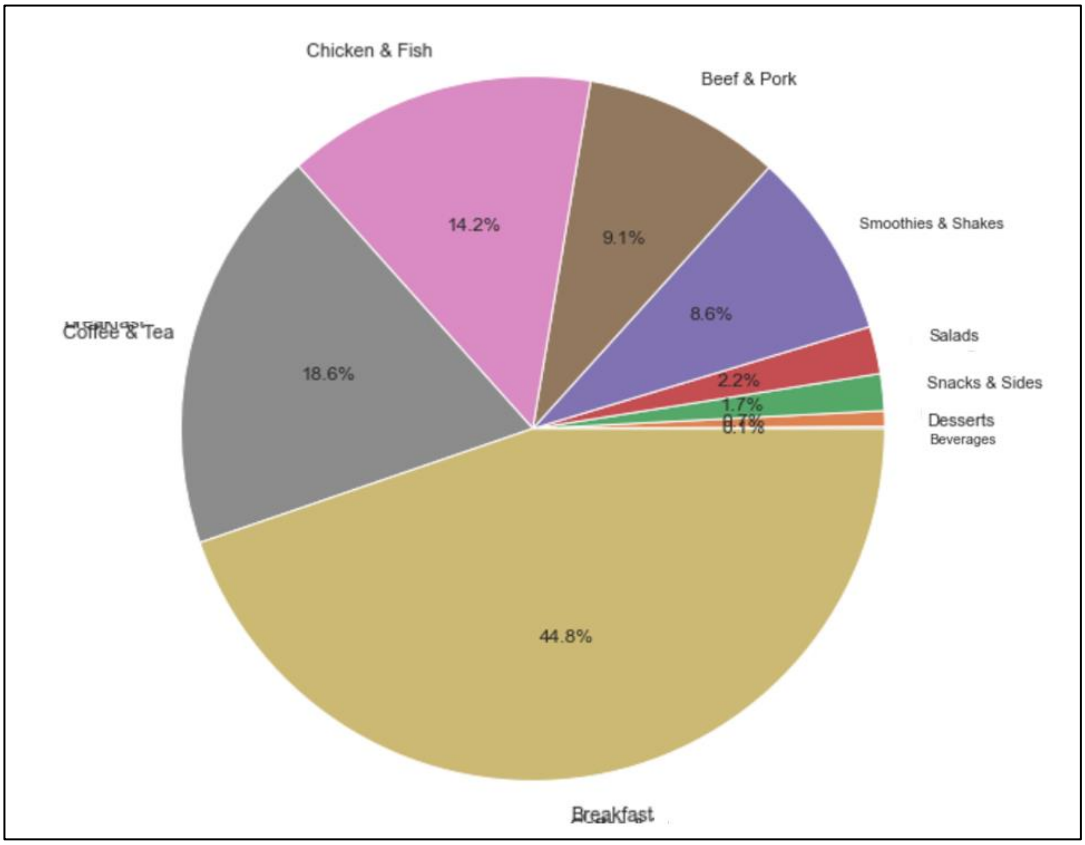
Plot 4: Boxplot with outliers for each variable

Q3. Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?

Breakfast (44.75%) contributes to maximum percentage of cholesterol in a diet (%daily value), with average breakfast cholesterol level of about '50%'.

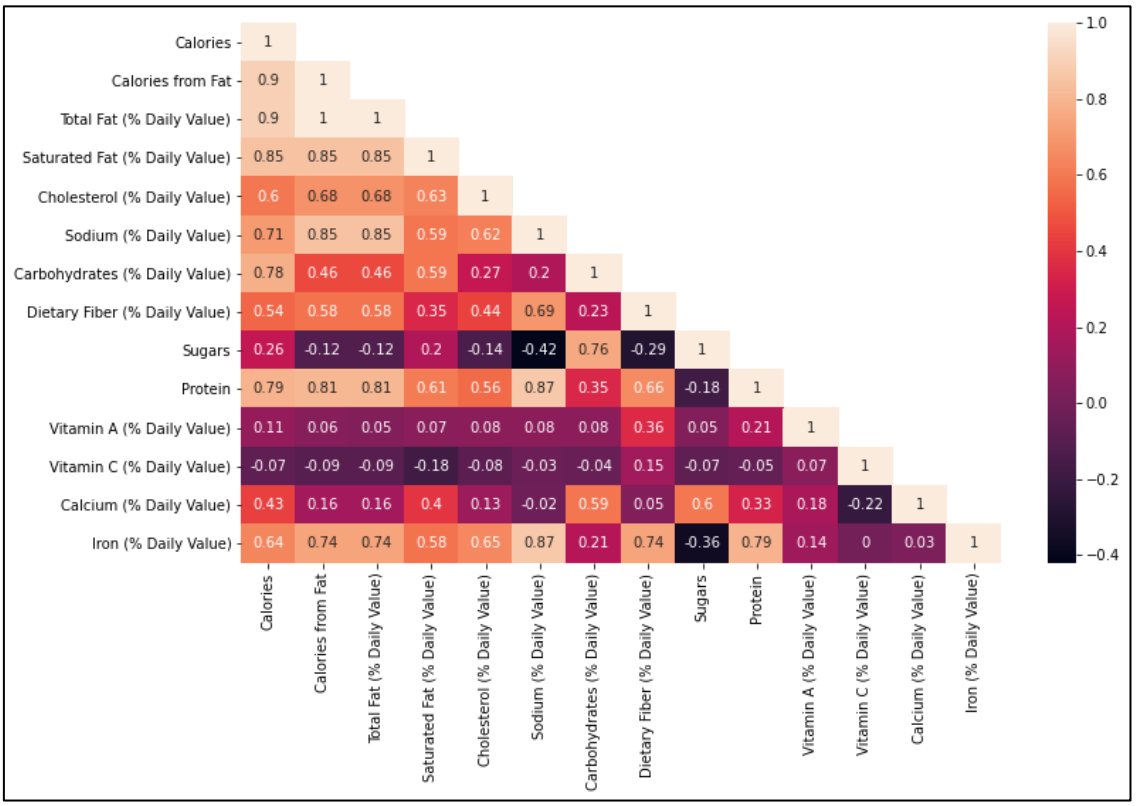
Beverages	0.10
Desserts	0.71
Snacks & Sides	1.69
Salads	2.17
Smoothies & Shakes	8.62
Beef & Pork	9.08
Chicken & Fish	14.24
Coffee & Tea	18.63
Breakfast	44.75

Plot 4:



Plot 5: Pie Chart showing percentage of cholesterol in a diet (%daily value) in each category

Q4. Which variables have the highest correlation? Plot them and find out the value?



Plot 5: Heatmap showing correlation between different variables

With the heatmap* given above, following can be established-

a) Calories have stronger relation with % daily values of Saturated Fats, Carbohydrate and Proteins. Hence, it can be said that food served by McDonalds derives calories from Saturated Fats, Carbohydrate and Proteins.
b) Sodium has stronger relation with % daily values of Total Fats, Protein and Iron. Hence, it may be that fatty and protein rich food has more sodium content in them.
c) Carbohydrates has stronger relation with % daily values sugars
d) Vitamins & Minerals generally have weak correlation with other variables of food nutrients

**the heatmap is constructed for %daily value variables only.*

Q5. Which item contributes maximum to the Sodium intake?
Chicken McNuggets contributes to most sodium (3600) and saturated fat intake (20).

```
1 sodium_total= df.groupby('Item')['Sodium'].sum().sort_values(ascending=False)
2 sodium_total
```

```
Item
Chicken McNuggets (40 piece)                                    3600
Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)      2290
Big Breakfast with Hotcakes (Large Biscuit)                     2260
Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit)    2170
Big Breakfast with Hotcakes (Regular Biscuit)                   2150
```

Q6. Which 4 food items contain the most amount of Saturated Fat?
The other food item high in saturated fat intake along with Chicken McNuggets are-

- McFlurry with M&M's Candies (Medium)
- Big Breakfast with Hotcakes (Large Biscuit)
- Chicken McNuggets (40 piece)
- Frappé Chocolate Chip (Large)

```
1 Sat_Fat_total= df.groupby('Item')['Saturated Fat'].sum().sort_values(ascending=False)
2 Sat_Fat_total
```

```
Item
McFlurry with M&M's Candies (Medium)          20.0
Big Breakfast with Hotcakes (Large Biscuit)   20.0
Chicken McNuggets (40 piece)                  20.0
Frappé Chocolate Chip (Large)                 20.0
Double Quarter Pounder with Cheese            19.0
```