Domain name:

- McDonald's food

Data Description:

- Based on the given data set, we will know that what kind of food iterms are most popular in mc donalds and what are the diffrent accepts for analysing the data.

Questions:

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

Import all the necessary libraries

```
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         sns.set_palette("deep")
         %matplotlib inline
         import seaborn as sns
         import os
         import scipy.stats as stats
```

Primary Analysis of the Dataset

In [2]:
```python
mcd = pd.read_csv("D:\\my files\\Mcdonald.csv")
mcd.head()
```

Out[2]:

	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	Diet Fi D Va
0	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13.0	20	5.0	25	0.0	...	31	10	4	
1	Breakfast	Egg White Delight	4.8 oz (135 g)	250	70	8.0	12	3.0	15	0.0	...	30	10	4	
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200	23.0	35	8.0	42	0.0	...	29	10	4	
3	Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)	450	250	28.0	43	10.0	52	0.0	...	30	10	4	
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	

5 rows × 24 columns

In [3]:
```python
mcd.describe()
```

Out[3]:

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohy
count	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	260.000000	...	260.
mean	368.269231	127.096154	14.165385	21.815385	6.007692	29.965385	0.203846	54.942308	18.392308	495.750000	...	47.
std	240.269886	127.875914	14.205998	21.885199	5.321873	26.639209	0.429133	87.269257	29.091653	577.026323	...	28.
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohy
25%	210.000000	20.000000	2.375000	3.750000	1.000000	4.750000	0.000000	5.000000	2.000000	107.500000	...	30.
50%	340.000000	100.000000	11.000000	17.000000	5.000000	24.000000	0.000000	35.000000	11.000000	190.000000	...	44.
75%	500.000000	200.000000	22.250000	35.000000	10.000000	48.000000	0.000000	65.000000	21.250000	865.000000	...	60.
max	1880.000000	1060.000000	118.000000	182.000000	20.000000	102.000000	2.500000	575.000000	192.000000	3600.000000	...	141.

8 rows × 21 columns

In [4]: `mcd.columns`

Out[4]:
```
Index(['Category', 'Item', '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)'],
      dtype='object')
```

- Total numbers of columns are 21 and list of columns are mention abow.

Additional information

In [5]: `mcd.shape`

Out[5]: `(260, 24)`

- total number of rows and columns are 260 rows and 24 columns.

In [6]: `mcd.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 24 columns):
```

```
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Category                       260 non-null    object
 1   Item                           260 non-null    object
 2   Serving Size                   260 non-null    object
 3   Calories                       260 non-null    int64
 4   Calories from Fat              260 non-null    int64
 5   Total Fat                      260 non-null    float64
 6   Total Fat (% Daily Value)      260 non-null    int64
 7   Saturated Fat                  260 non-null    float64
 8   Saturated Fat (% Daily Value)  260 non-null    int64
 9   Trans Fat                      260 non-null    float64
 10  Cholesterol                    260 non-null    int64
 11  Cholesterol (% Daily Value)    260 non-null    int64
 12  Sodium                         260 non-null    int64
 13  Sodium (% Daily Value)         260 non-null    int64
 14  Carbohydrates                  260 non-null    int64
 15  Carbohydrates (% Daily Value)  260 non-null    int64
 16  Dietary Fiber                  260 non-null    int64
 17  Dietary Fiber (% Daily Value)  260 non-null    int64
 18  Sugars                         260 non-null    int64
 19  Protein                        260 non-null    int64
 20  Vitamin A (% Daily Value)      260 non-null    int64
 21  Vitamin C (% Daily Value)      260 non-null    int64
 22  Calcium (% Daily Value)        260 non-null    int64
 23  Iron (% Daily Value)           260 non-null    int64
dtypes: float64(3), int64(18), object(3)
memory usage: 45.8+ KB
```

- Data types:- Integer, Float and Object.

- Missing Values:- no missing values available in data set.

- Number of total values:- 260 all

In [7]: `mcd['Category'].value_counts()`

Out[7]:
```
Coffee & Tea        95
Breakfast           42
Smoothies & Shakes  28
Beverages           27
Chicken & Fish      27
Beef & Pork         15
Snacks & Sides      13
Desserts             7
Salads               6
Name: Category, dtype: int64
```

In [8]: `mcd['Category'].nunique()`

Out[8]: 9

- In data set 9 Number of unique values in category.

In [9]: `mcd['Item'].value_counts()`

Out[9]:
```
Premium Grilled Chicken Classic Sandwich             1
Premium McWrap Chicken Sweet Chili (Crispy Chicken)  1
Bacon Clubhouse Grilled Chicken Sandwich             1
Latte with Sugar Free French Vanilla Syrup (Small)   1
Coffee (Large)                                       1
                                                    ..
Steak & Egg McMuffin                                 1
Caramel Iced Coffee (Medium)                         1
Coca-Cola Classic (Medium)                           1
Sprite (Small)                                       1
Caramel Latte (Medium)                               1
Name: Item, Length: 260, dtype: int64
```

In [10]: `mcd['Item'].nunique()`

Out[10]: 260

- Number of 260 unique items are available in data set .

In [11]: `mcd['Serving Size'].value_counts()`

Out[11]:
```
16 fl oz cup      45
12 fl oz cup      38
22 fl oz cup      20
20 fl oz cup      16
21 fl oz cup       7
                  ..
10.3 oz (291 g)    1
10.7 oz (304 g)    1
6.2 oz (177 g)     1
6.4 oz (182 g)     1
4.3 oz (121 g)     1
Name: Serving Size, Length: 107, dtype: int64
```

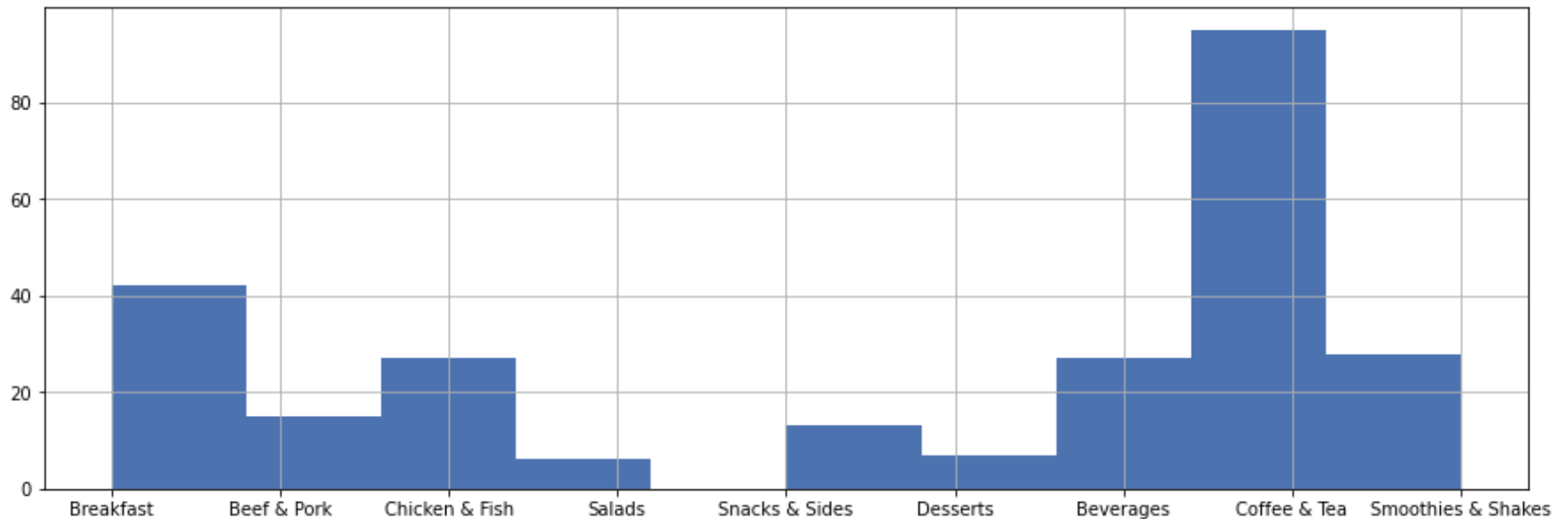In [12]: `mcd['Serving Size'].nunique()`

Out[12]: 107

- In data set 107 Number of unique values in Serving size.

Q1. Plot graphically which food categories have the highest and lowest varieties.

```
In [13]:   plt.figure(figsize=(15,5))
           mcd["Category"].hist()
```

Out[13]:   <AxesSubplot:>

As per analysis

- "coffee & tea" have highest varieties &
- "salads" & "Desserts " have lowest varieties.

Q2.Which all variables have an outlier?

```
In [14]:   mcd.columns
```
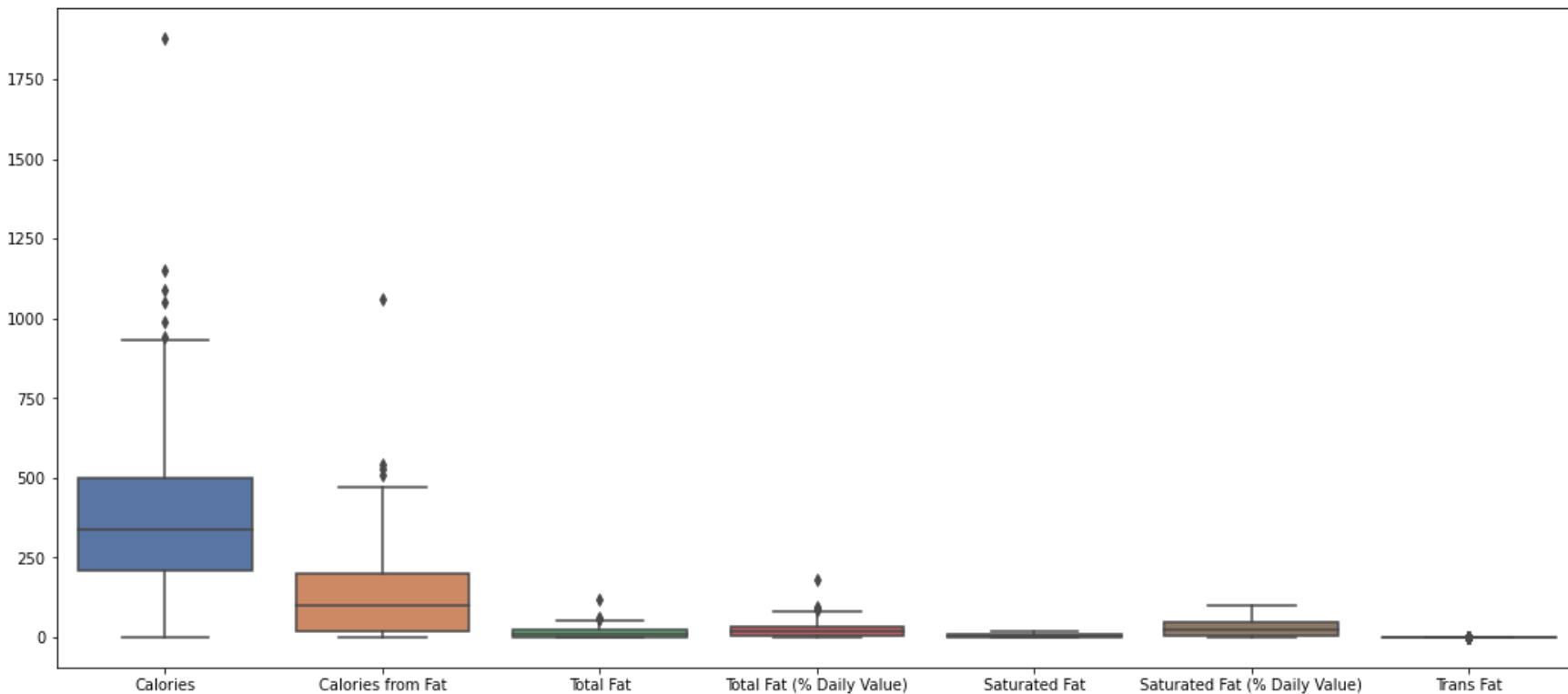
Out[14]:   Index(['Category', 'Item', 'Serving Size', 'Calories', 'Calories from Fat',
                  'Total Fat', 'Total Fat (% Daily Value)', 'Saturated Fat',
                  'Saturated Fat (% Daily Value)', 'Trans Fat', 'Cholesterol',
                  'Cholesterol (% Daily Value)', 'Sodium', '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)'],
dtype='object')
```

In [15]:
```
plt.figure(figsize=(18,8))

sns.boxplot(data=mcd[['Calories', 'Calories from Fat', 'Total Fat', 'Total Fat (% Daily Value)', 'Saturated Fat', 'Satura
```
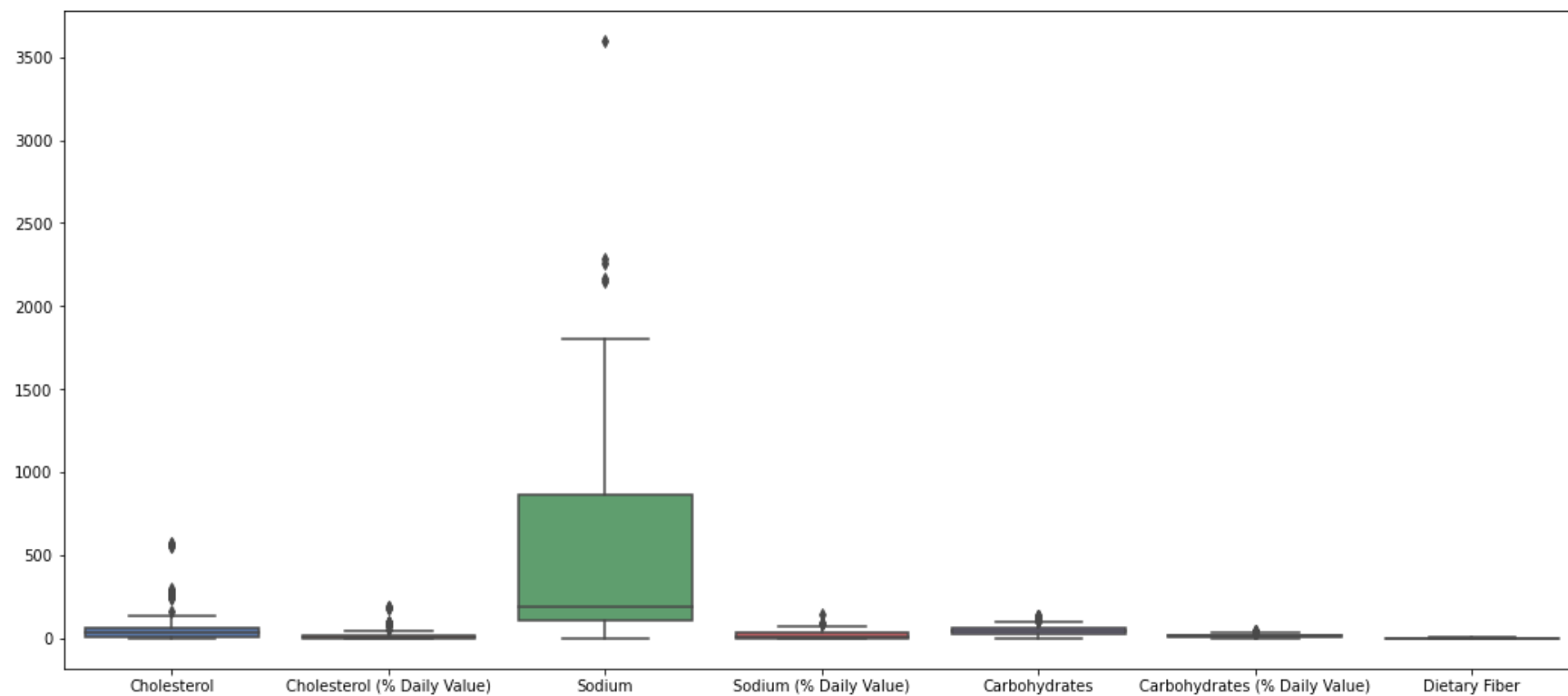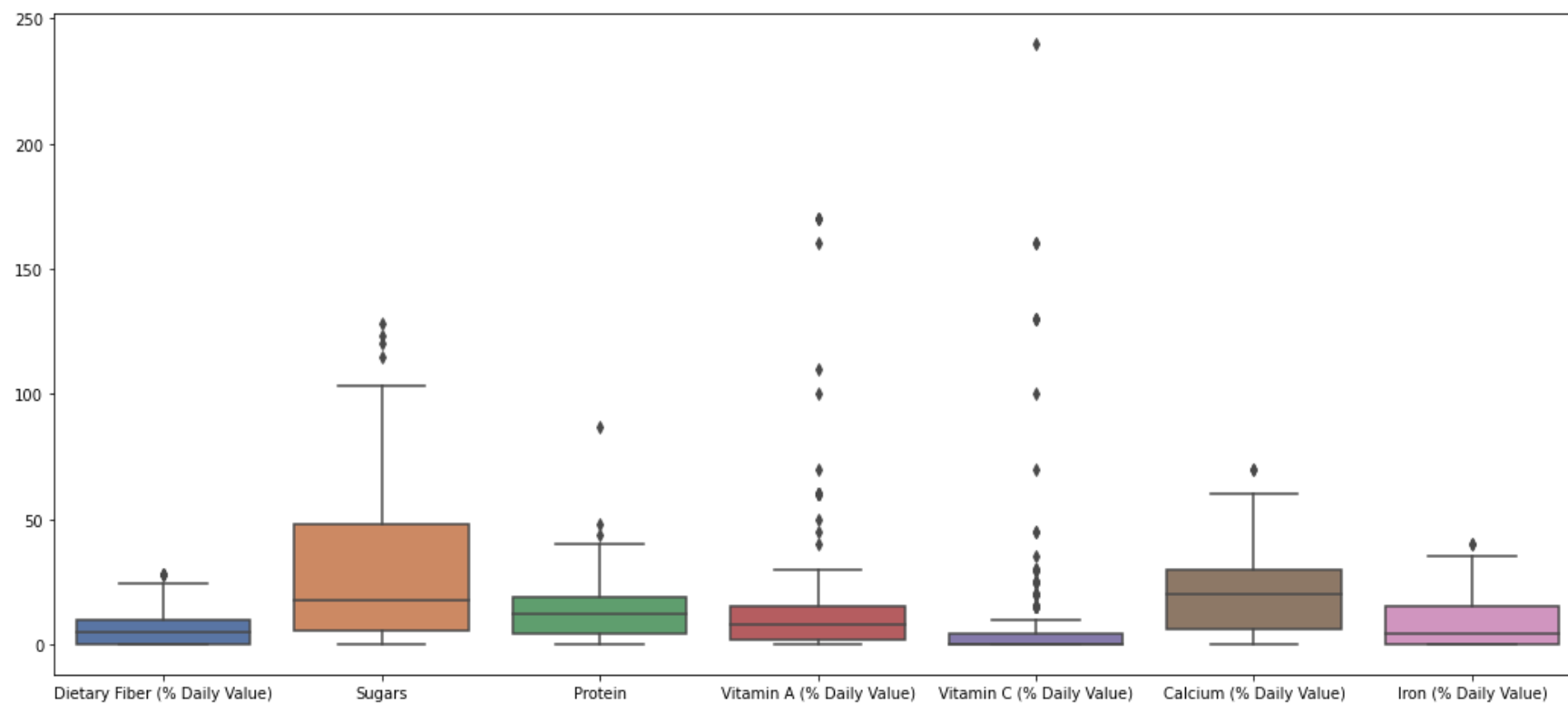
Out[15]: <AxesSubplot:>

In [16]:
```
plt.figure(figsize=(18,8))

sns.boxplot(data=mcd[['Cholesterol','Cholesterol (% Daily Value)', 'Sodium', 'Sodium (% Daily Value)','Carbohydrates', 'C
```

Out[16]: <AxesSubplot:>

```
In [17]:  plt.figure(figsize=(18,8))

          sns.boxplot(data=mcd[['Dietary Fiber (% Daily Value)','Sugars','Protein','Vitamin A (% Daily Value)', 'Vitamin C (% Daily
```

Out[17]: <AxesSubplot:>

The list of columns with outlier :-

- Calories
- Calories from Fat
- Total fat
- Saturated Fat (% Daily Value)
- Cholestrol
- Cholestrol (% Daily Value)
- Calories from Fat
- Total Fat
- Total Fat (% Daily Value)
- Sodium (% Daily Value)
- Carbohydrates
- Carbohydrates (% Daily Value)
- Dietary Fiber (% Daily Value)
- Sugars

- Protien
- Vitamin A (% Daily Value)
- Vitamin C (% Daily Value)
- Calcium (% Daily Value)
- Iron (% Daily Value)

List of columns without outlier :-

- Saturated Fat
- Trans Fat
- Sodium
- Dietary Fiber

Q3. Which variables have the highest correlation? Plot them and find out the value?

```
In [18]:  corr=mcd.corr()
          corr
```

Out[18]:

|	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohydrate
Calories	1.000000	0.904588	0.904409	0.904123	0.845564	0.847631	0.522441	0.596399	0.595208	0.712309	...	0.78153
Calories from Fat	0.904588	1.000000	0.999663	0.999725	0.847008	0.849592	0.433686	0.682161	0.681607	0.846624	...	0.46167
Total Fat	0.904409	0.999663	1.000000	0.999765	0.846707	0.849293	0.431453	0.680547	0.680000	0.846158	...	0.46121
Total Fat (% Daily Value)	0.904123	0.999725	0.999765	1.000000	0.847379	0.849973	0.433016	0.680940	0.680378	0.846728	...	0.46051
Saturated Fat	0.845564	0.847008	0.846707	0.847379	1.000000	0.999279	0.620611	0.631210	0.630334	0.584075	...	0.59126
Saturated Fat (% Daily Value)	0.847631	0.849592	0.849293	0.849973	0.999279	1.000000	0.620210	0.633603	0.632712	0.588694	...	0.59132
Trans Fat	0.522441	0.433686	0.431453	0.433016	0.620611	0.620210	1.000000	0.253935	0.251502	0.187580	...	0.46325
Cholesterol	0.596399	0.682161	0.680547	0.680940	0.631210	0.633603	0.253935	1.000000	0.999855	0.624362	...	0.27097

Mc' Donald's

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholesterol	Cholesterol (% Daily Value)	Sodium	...	Carbohydrate
Cholesterol (% Daily Value)	0.595208	0.681607	0.680000	0.680378	0.630334	0.632712	0.251502	0.999855	1.000000	0.623320	...	0.26930
Sodium	0.712309	0.846624	0.846158	0.846728	0.584075	0.588694	0.187580	0.624362	0.623320	1.000000	...	0.20079
Sodium (% Daily Value)	0.713415	0.847276	0.846780	0.847368	0.585323	0.589958	0.188339	0.624743	0.623720	0.999929	...	0.20242
Carbohydrates	0.781539	0.461672	0.461213	0.460516	0.591261	0.591322	0.463250	0.270977	0.269300	0.200796	...	1.00000
Carbohydrates (% Daily Value)	0.781242	0.461463	0.461005	0.460298	0.591743	0.591655	0.462891	0.272662	0.270992	0.201032	...	0.99962
Dietary Fiber	0.538894	0.581274	0.580837	0.580592	0.351818	0.356831	0.054918	0.435575	0.434940	0.694389	...	0.22457
Dietary Fiber (% Daily Value)	0.540014	0.575621	0.575206	0.575033	0.347152	0.351797	0.058301	0.440266	0.439814	0.689995	...	0.22825
Sugars	0.259598	-0.115285	-0.115446	-0.115761	0.197734	0.195928	0.334756	-0.135518	-0.136459	-0.426536	...	0.76236
Protein	0.787847	0.807913	0.807773	0.807922	0.603028	0.606581	0.388249	0.561561	0.560957	0.869802	...	0.35212
Vitamin A (% Daily Value)	0.108844	0.056731	0.054434	0.054038	0.064972	0.065376	0.075833	0.080239	0.080059	0.083068	...	0.08380
Vitamin C (% Daily Value)	-0.068747	-0.087331	-0.089354	-0.089353	-0.179672	-0.178059	-0.076612	-0.082978	-0.083315	-0.030769	...	-0.03472
Calcium (% Daily Value)	0.428426	0.161034	0.162860	0.162031	0.403311	0.401139	0.385331	0.132077	0.132382	-0.024074	...	0.58969
Iron (% Daily Value)	0.643552	0.735894	0.734685	0.735478	0.578062	0.580488	0.325476	0.655000	0.653167	0.871593	...	0.21024

21 rows × 21 columns

In [19]:
```python
plt.figure(figsize=(20,10))
sns.heatmap(corr,annot=True)
```

```
Out[19]: <AxesSubplot:>
```
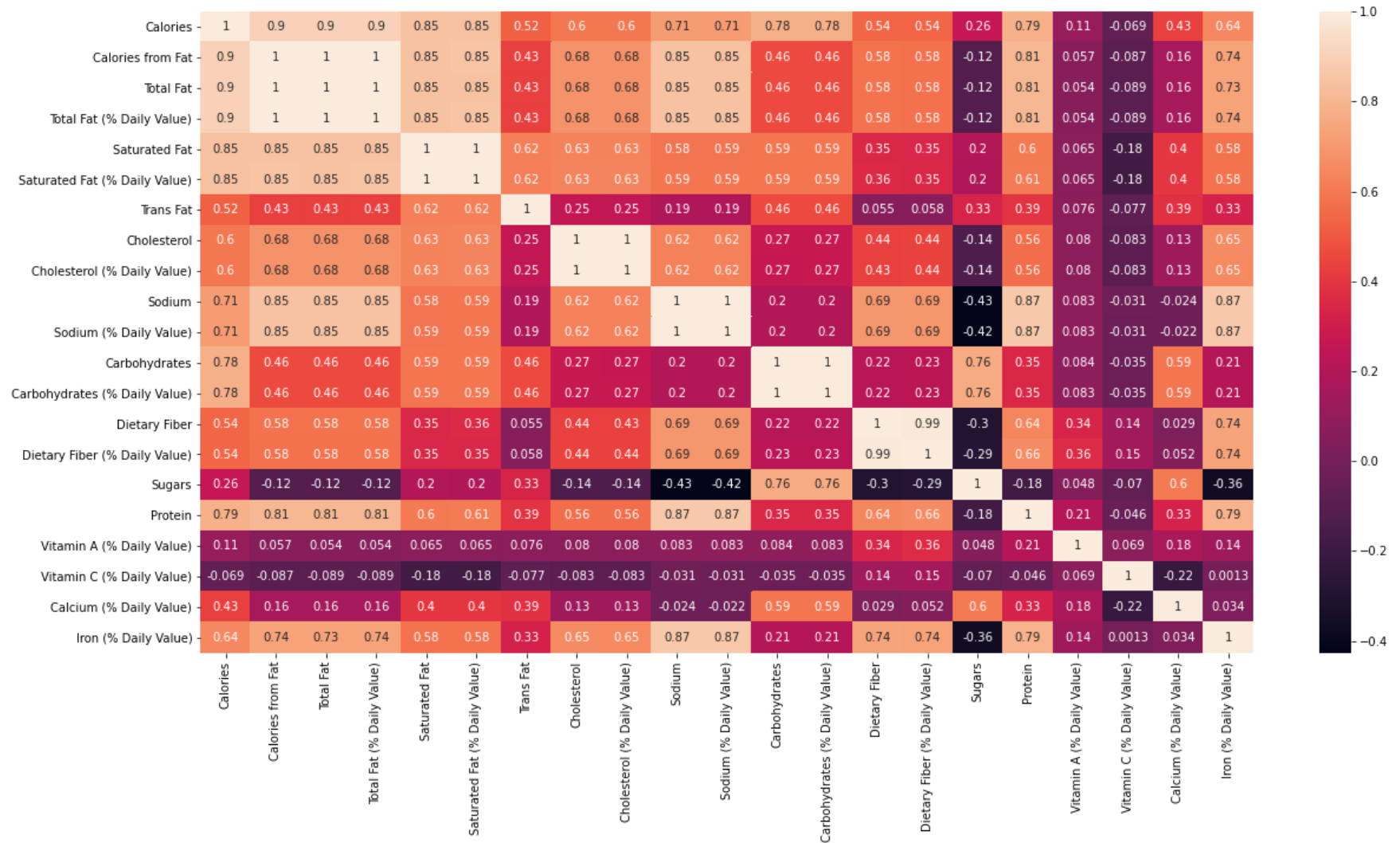

Q4. Which category contributes to the maximum % of Cholesterol in a diet (% daily value)?

```python
In [20]: macd1= pd.pivot_table(mcd, 'Cholesterol (% Daily Value)', index=['Category'])

me = macd1.sort_values(('Cholesterol (% Daily Value)'), ascending=False)

me.head(10)
```

Out[20]:

	Cholesterol (% Daily Value)
Category	
Breakfast	50.952381
Beef & Pork	28.933333
Chicken & Fish	25.222222
Salads	17.333333
Smoothies & Shakes	14.714286
Coffee & Tea	9.378947
Snacks & Sides	6.230769
Desserts	4.857143
Beverages	0.185185

- The maximum persantage of Cholesterol is in Breakfast which have (50.95%) almost 51% of Cholestrol.

Q5. Which item contributes maximum to the Sodium intake?

In [21]:
```python
mcd2 = pd.pivot_table(mcd, 'Sodium', index=['Item'])

me = mcd2.sort_values(('Sodium'), ascending=False)

me.head(10)
```

Out[21]:

	Sodium
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

	Sodium
Item	
Chicken McNuggets (20 piece)	1800
Bacon Clubhouse Crispy Chicken Sandwich	1720
Big Breakfast with Egg Whites (Large Biscuit)	1700
Big Breakfast (Large Biscuit)	1680
Big Breakfast with Egg Whites (Regular Biscuit)	1590

- The maximum Sodium intake item is Chicken McNuggets (40 piece) with 3600 sodium intake value.

Q6. Which 4 food items contain the most amount of Saturated Fat?

In [22]:
```python
mcd3 = pd.pivot_table(mcd, 'Saturated Fat', index=['Item'])

me = mcd3.sort_values(('Saturated Fat'), ascending=False)

me.head(4)
```

Out[22]:

	Saturated Fat
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

4 food items contain the most amount of Saturated Fat are :-

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

Thank you