In [10]:
```
import numpy as np
import pandas as pd
from numpy.random import randn
from scipy.stats import f
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import seaborn as sborn
import scipy.stats as stats
```

In [11]:
```
df =  pd.read_csv('C:\\Users\\milon\\Mcdonald .csv')
data = df
data
```

Out[11]:

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

	Category	Item	Serving Size	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat
255	Smoothies & Shakes	McFlurry with Oreo Cookies (Small)	10.1 oz (285 g)	510	150	17.0	26	9.0	44	0.5
256	Smoothies & Shakes	McFlurry with Oreo Cookies (Medium)	13.4 oz (381 g)	690	200	23.0	35	12.0	58	1.0
257	Smoothies & Shakes	McFlurry with Oreo Cookies (Snack)	6.7 oz (190 g)	340	100	11.0	17	6.0	29	0.0
258	Smoothies & Shakes	McFlurry with Reese's Peanut Butter Cups (Medium)	14.2 oz (403 g)	810	290	32.0	50	15.0	76	1.0
259	Smoothies & Shakes	McFlurry with Reese's Peanut Butter Cups (Snack)	7.1 oz (202 g)	410	150	16.0	25	8.0	38	0.0

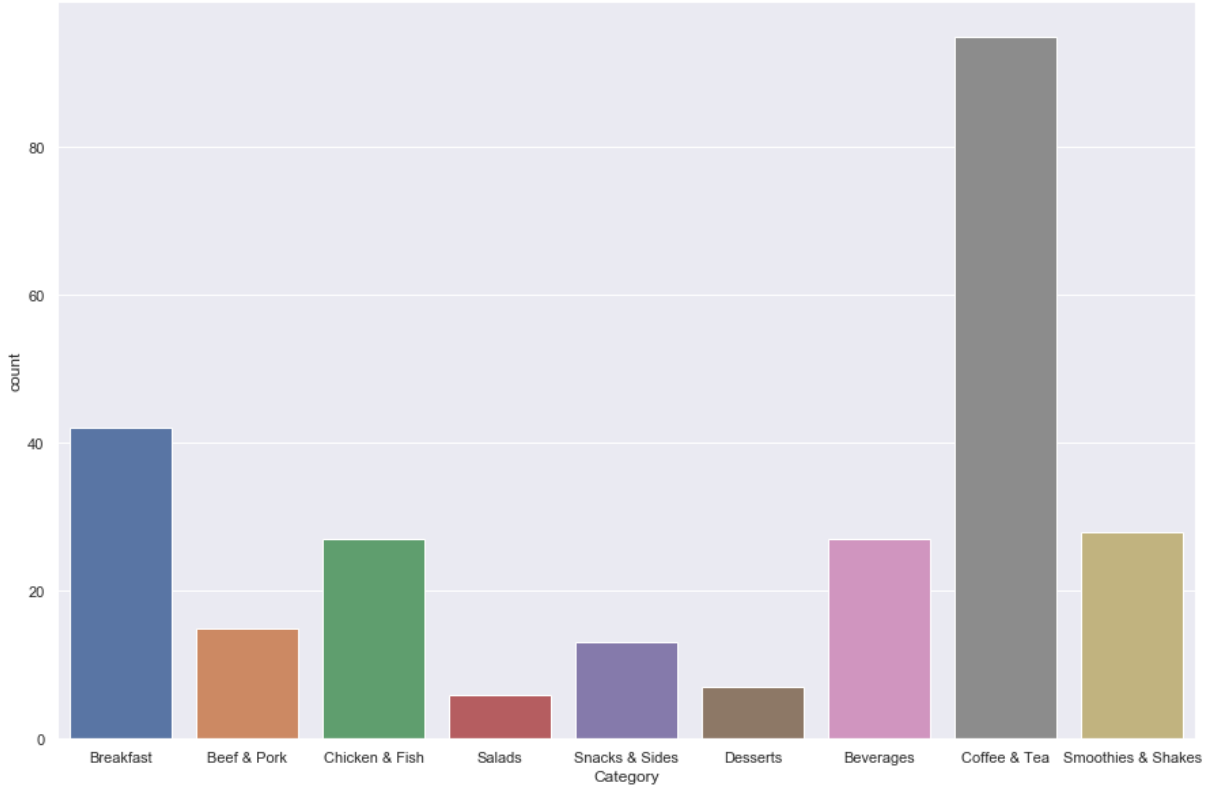
260 rows × 24 columns

In [12]:
```
df['Category'].unique() # total varities of catagory
df['Category'].value_counts() # counting of unique catagories
```

Out[12]:
```
Coffee & Tea          95
Breakfast             42
Smoothies & Shakes    28
Chicken & Fish        27
Beverages             27
Beef & Pork           15
Snacks & Sides        13
Desserts               7
Salads                 6
Name: Category, dtype: int64
```

In [154]:
```
plt.figure(figsize=(15,10))
sns.countplot(x="Category",data=data)  # count plot
```

Out[154]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2a34b461dc8>
```

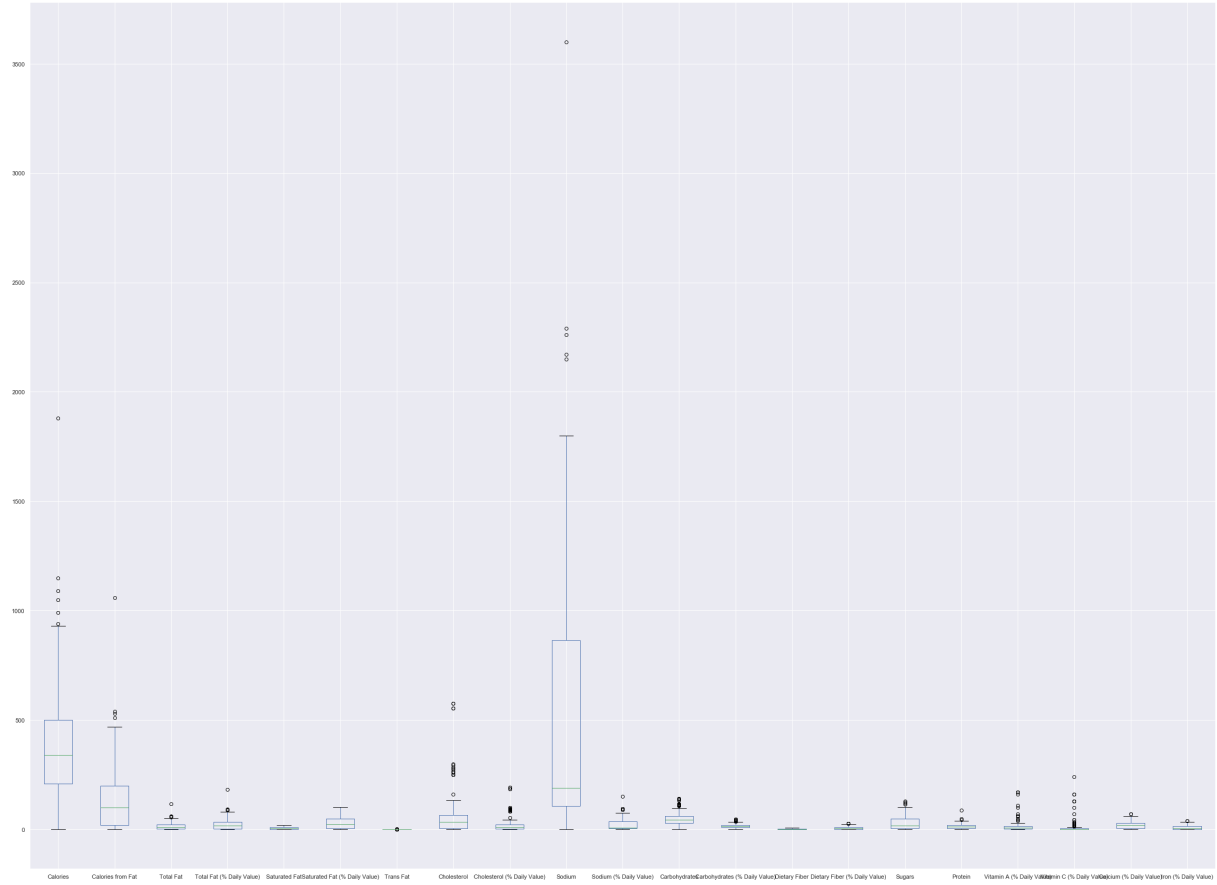


In [155]:
```
Q1 = data.quantile(0.25)
Q3 = data.quantile(0.75)
IQR = Q3 - Q1
lb = Q1 -(1.5 * IQR)
ub = Q3 +(1.5 * IQR)
gh=((data< (Q1 - 1.5 * IQR)) | (data> (Q3 + 1.5 * IQR)))
gh
gh[gh == 1].count()# list of variables with outliers and number of outl
iers in it
```

Out[155]:
```
Calcium (% Daily Value)           2
Calories                          6
Calories from Fat                 4
Carbohydrates                    17
Carbohydrates (% Daily Value)    16
Category                          0
Cholesterol                      18
Cholesterol (% Daily Value)      18
Dietary Fiber                     0
Dietary Fiber (% Daily Value)     4
Iron (% Daily Value)              2
Item                              0
Protein                           3
Saturated Fat                     0
Saturated Fat (% Daily Value)     0
Serving Size                      0
Sodium                            5
Sodium (% Daily Value)            5
Sugars                            4
Total Fat                         4
Total Fat (% Daily Value)         4
Trans Fat                        56
Vitamin A (% Daily Value)        17
Vitamin C (% Daily Value)        46
dtype: int64
```

In [103]:
```
dn=data
sborn.set(font_scale=1)
```

```
plt.figure(figsize=(40,30))
_, fig = pd.DataFrame.boxplot(dn, return_type='both')
outliers = [flier.get_ydata() for flier in bp["fliers"]]
out_liers = [i.tolist() for i in outliers]
```



In [3]:
```
corr = data.corr()
```

In [159]:
```
cor = data[['Category','Item', 'Serving Size','Calories','Calories from
 Fat','Total Fat','Total Fat (% Daily Value)','Saturated Fat','Saturate
d Fat (% Daily Value)',
 'Trans Fat','Cholesterol','Cholesterol (% Daily Value)','Sodium','Sodi
um (% Daily Value)','Carbohydrates','Carbohydrates (% Daily Value)',
```

```
 'Dietary Fiber','Dietary Fiber (% Daily Value)','Sugars','Protein','Vi
tamin A (% Daily Value)','Vitamin C (% Daily Value)','Calcium (% Daily
 Value)','Iron (% Daily Value)']].corr()
cor
```

Out[159]:

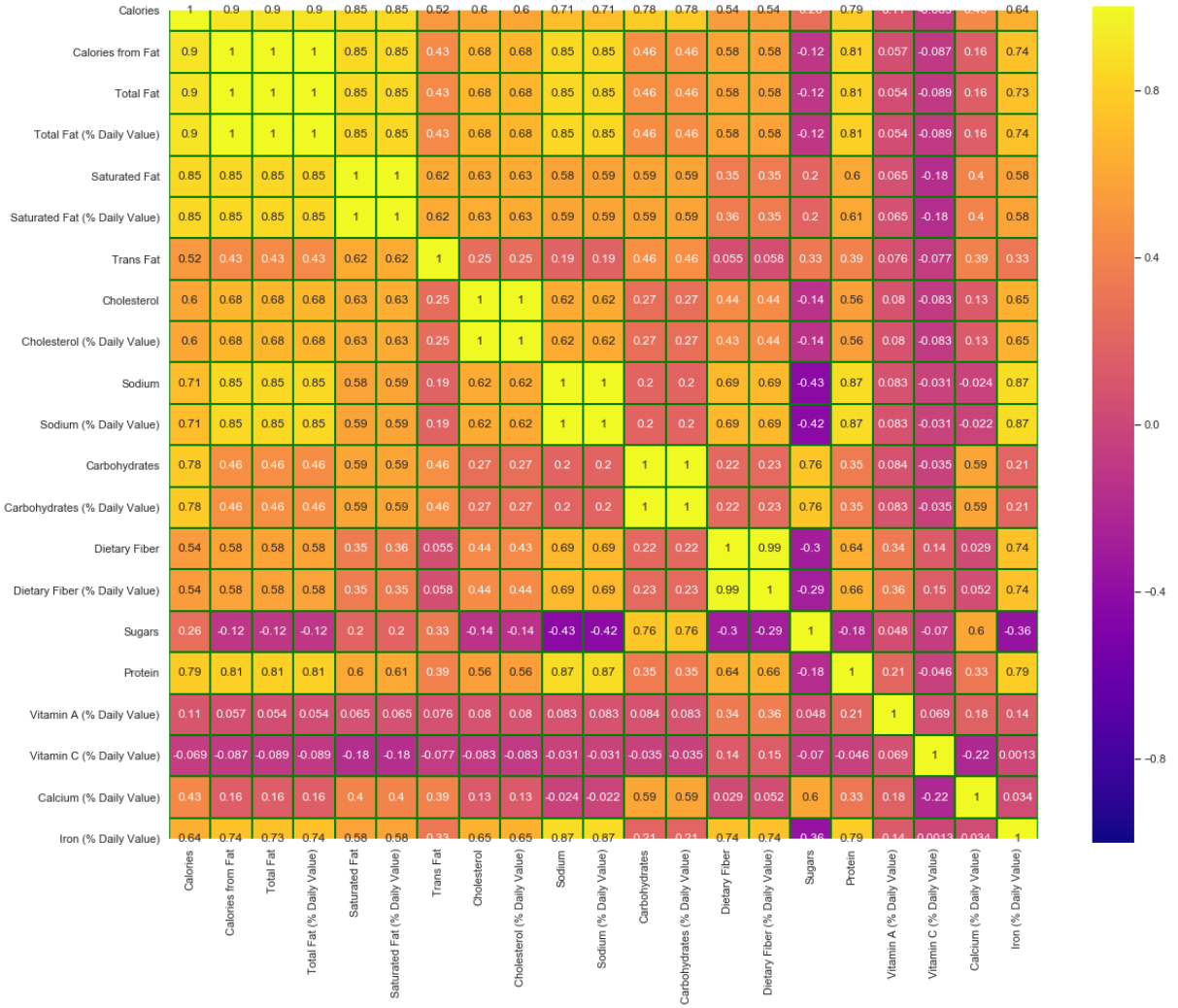
	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Choles
Calories	1.000000	0.904588	0.904409	0.904123	0.845564	0.847631	0.522441	0.59
Calories from Fat	0.904588	1.000000	0.999663	0.999725	0.847008	0.849592	0.433686	0.68
Total Fat	0.904409	0.999663	1.000000	0.999765	0.846707	0.849293	0.431453	0.68
Total Fat (% Daily Value)	0.904123	0.999725	0.999765	1.000000	0.847379	0.849973	0.433016	0.68
Saturated Fat	0.845564	0.847008	0.846707	0.847379	1.000000	0.999279	0.620611	0.63
Saturated Fat (% Daily Value)	0.847631	0.849592	0.849293	0.849973	0.999279	1.000000	0.620210	0.63
Trans Fat	0.522441	0.433686	0.431453	0.433016	0.620611	0.620210	1.000000	0.25
Cholesterol	0.596399	0.682161	0.680547	0.680940	0.631210	0.633603	0.253935	1.00
Cholesterol (% Daily Value)	0.595208	0.681607	0.680000	0.680378	0.630334	0.632712	0.251502	0.99
Sodium	0.712309	0.846624	0.846158	0.846728	0.584075	0.588694	0.187580	0.62
Sodium (% Daily Value)	0.713415	0.847276	0.846780	0.847368	0.585323	0.589958	0.188339	0.62
Carbohydrates	0.781539	0.461672	0.461213	0.460516	0.591261	0.591322	0.463250	0.27
Carbohydrates (% Daily Value)	0.781242	0.461463	0.461005	0.460298	0.591743	0.591655	0.462891	0.27
Dietary Fiber	0.538894	0.581274	0.580837	0.580592	0.351818	0.356831	0.054918	0.43
Dietary Fiber (% Daily Value)	0.540014	0.575621	0.575206	0.575033	0.347152	0.351797	0.058301	0.44

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Choles
Sugars	0.259598	-0.115285	-0.115446	-0.115761	0.197734	0.195928	0.334756	-0.13
Protein	0.787847	0.807913	0.807773	0.807922	0.603028	0.606581	0.388249	0.56
Vitamin A (% Daily Value)	0.108844	0.056731	0.054434	0.054038	0.064972	0.065376	0.075833	0.08
Vitamin C (% Daily Value)	-0.068747	-0.087331	-0.089354	-0.089353	-0.179672	-0.178059	-0.076612	-0.08
Calcium (% Daily Value)	0.428426	0.161034	0.162860	0.162031	0.403311	0.401139	0.385331	0.13
Iron (% Daily Value)	0.643552	0.735894	0.734685	0.735478	0.578062	0.580488	0.325476	0.65

21 rows × 21 columns

In [160]:
```
plt.figure(figsize=(19, 15))
sborn.set(font_scale=1)
sns.heatmap(cor,center=False,robust=False,annot=True, cmap='plasma',vmi
n=-1,vmax=1,fmt='.2g',
    annot_kws=None,
    linewidths=1,
    linecolor='green',
    cbar=True,
    cbar_kws=None,
    cbar_ax=None,
    square=True,
    xticklabels='auto',
    yticklabels='auto',
    mask=None,
    ax=None)
```

Out[160]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2a348fd3e08>
```



In [161]:
```
data['Category'].unique()
```

Out[161]:
```
array(['Breakfast', 'Beef & Pork', 'Chicken & Fish', 'Salads',
       'Snacks & Sides', 'Desserts', 'Beverages', 'Coffee & Tea',
       'Smoothies & Shakes'], dtype=object)
```

In [162]:
```
mp = data[['Category','Cholesterol (% Daily Value)']]
```

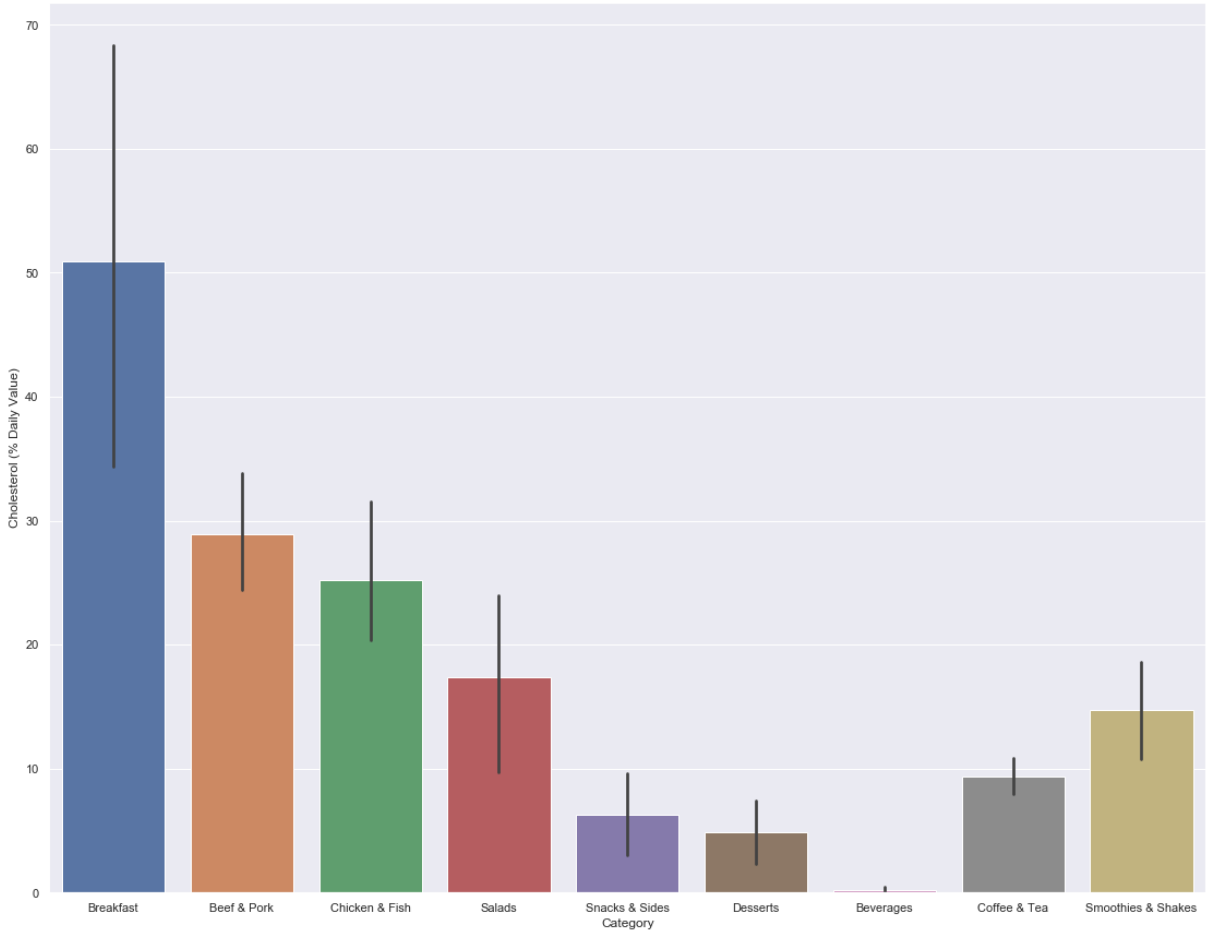
```
mp
df2 = mp.groupby(['Category']).sum()
df2
df2.sort_values('Cholesterol (% Daily Value)',ascending=False)
```

Out[162]:

Category	Cholesterol (% Daily Value)
Breakfast	2140
Coffee & Tea	891
Chicken & Fish	681
Beef & Pork	434
Smoothies & Shakes	412
Salads	104
Snacks & Sides	81
Desserts	34
Beverages	5

In [163]:
```
plt.figure(figsize=(19, 15))
sns.barplot(x="Category", y="Cholesterol (% Daily Value)", data=data)
```

Out[163]: <matplotlib.axes._subplots.AxesSubplot at 0x2a345e74f08>



In [164]:
```
mq = data[['Item','Sodium']]
mq
df3 = mq.groupby(['Item']).sum()
df3
fg = df3.sort_values('Sodium',ascending=False)
fg.head(10)
```

Out[164]:

Item	Sodium

Item	Sodium
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
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

In [165]:
```
mn = data[['Item','Saturated Fat']]
mn
df4 = mn.groupby(['Item']).sum()
df4
gh = df4.sort_values('Saturated Fat',ascending=False)
gh.head(4)
```

Out[165]:

Item	Saturated Fat
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