Food_Nutrition_Analysis

April 18, 2023

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
sns.set(font_scale=1.5)
sns.set(rc={"figure.figsize":(40, 10)})
```

0.1 Read in data

[2]:
```
# Prepare the data. This script fixes the ABBREV.csv file and pulls the foods␣
 ↪into their respective food groups
%run data_prep.py
```

[26]:
```
averages_list = [water_average, meat_average, canned_food_average,␣
 ↪frozen_food_average, seafoods_average, seasonings_average, dessert_average,␣
 ↪fruits_average, cheeses_average, milks_average, cereals_average,␣
 ↪babyfoods_average, beans_average, rice_average, pasta_average,␣
 ↪fast_food_average, pizza_average, school_lunches_average, bread_average,␣
 ↪fish_average, na_beverages_average, alcoholic_beverages_average,␣
 ↪dressings_average, oils_average, butter_average, gravy_average,␣
 ↪veggies_average, nuts_average, snacks_average]

# combine all the dataframes into one using concat
all_averages = pd.concat(averages_list, ignore_index=True)

all_averages = all_averages.sort_values(by='Health_Score', ascending=False)
all_averages['Name'] = all_averages['Name'].str.replace('_Average', '')
all_averages['Name'] = all_averages['Name'].str.replace('_average', '')
all_averages = all_averages.drop('id', axis=1)
all_averages
```

[26]:

	Name	Calories	Protein_(mg)	Carbs_(mg)	Fiber_(mg)	\
0	water	0.33	0.00	43.33	0.00	
1	meat	188.37	22335.25	1480.18	74.65	
19	fish	173.93	18668.08	3459.69	335.14	

4	sea_foods	115.00	14444.21	6506.58	940.58
26	veggies	70.75	2258.19	10116.72	2330.41
2	canned_food	67.94	2865.58	9277.73	1315.06
12	beans	139.41	9366.28	22463.09	7245.93
7	fruits	85.80	1535.27	19218.96	2201.82
20	na_beverages	92.27	2769.25	15503.54	618.89
9	milks	127.62	8921.91	14776.17	70.21
8	cheeses	287.84	21502.70	5339.21	134.48
3	frozen_foods	202.28	10628.27	18796.29	2448.46
21	alcoholic_beverages	129.80	1025.27	7737.30	127.08
11	babyfoods	147.75	4015.73	21141.52	894.10
22	dressings/condiments	217.70	2334.84	17902.79	1025.41
14	pasta	144.96	5615.76	24432.64	2256.03
25	gravy	134.15	4158.97	20772.56	1148.00
15	fast_food	251.40	12195.61	22419.04	1435.52
17	school_lunch	249.15	13523.08	27435.38	3430.77
27	nuts	433.19	12127.14	28429.21	7612.24
16	pizza	265.20	12013.45	29713.10	2668.97
24	butter	566.76	15166.55	22978.97	4634.48
23	oils	789.23	1993.65	3052.29	164.89
5	seasonings	251.84	8406.40	50282.00	21264.00
18	bread	271.84	9474.18	48594.48	5084.13
13	rice	254.88	5548.54	53961.67	1917.50
6	dessert	387.99	5355.54	57653.35	2654.16
10	cereals	345.15	8340.40	74043.80	7722.29
28	snacks	459.40	6928.56	65323.22	3893.62

	Sugar_(mg)	Calcium_(mg)	Sodium_(mg)	Total_Fat_(mg)	Potassium_(mg)	\
0	0.00	5.00	5.67	0.00	1.33	
1	615.30	17.99	177.41	7231.17	294.02	
19	485.31	43.57	330.19	5847.88	345.68	
4	673.11	54.70	486.53	1573.08	282.22	
26	5298.86	59.56	138.52	474.95	286.50	
2	2495.42	17.50	455.42	1617.01	231.81	
12	2188.75	62.24	139.57	900.38	477.95	
7	13534.41	34.25	33.19	581.96	233.80	
20	11182.85	57.69	98.32	750.12	233.84	
9	15879.44	300.55	175.49	3290.34	406.04	
8	1699.04	581.78	754.25	17747.44	145.93	
3	4111.21	78.77	370.55	7029.89	233.85	
21	7724.59	12.33	25.35	598.46	82.46	
11	13834.81	135.63	54.27	4103.28	192.32	
22	10872.00	41.66	895.73	7594.59	152.72	
14	3796.94	41.87	305.26	1971.51	197.80	
25	4053.18	62.21	1903.54	4530.85	207.40	
15	4849.23	98.09	549.37	8671.48	211.08	
17	4965.38	165.00	461.46	6113.62	309.31	

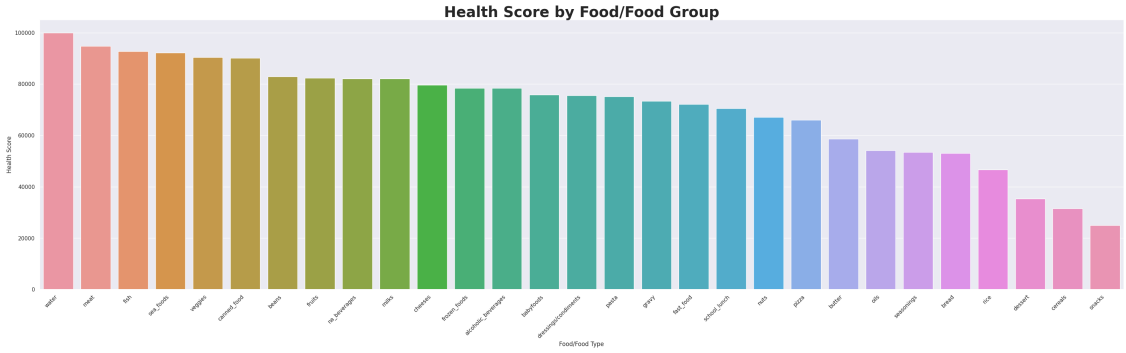
27	6714.17	81.38	107.70	22034.63	581.07
16	4031.22	178.46	584.67	7849.79	212.55
24	9829.55	130.31	277.41	35011.24	441.07
23	189.23	9.59	120.05	57971.22	72.05
5	3622.63	597.52	2752.40	4744.64	1071.35
18	5071.90	89.91	460.52	2531.25	188.97
13	858.40	19.44	83.31	861.80	145.04
6	25868.63	97.09	374.59	10904.00	200.54
10	20624.17	159.30	382.33	2243.61	274.86
28	34553.79	132.37	299.48	13444.63	335.94

	Water_(mg)	Vitamins_(mg)	Unhealthy_Nutrients	Healthy_Nutrients	\
0	99943.33	0.00	0.00	99956.67	
1	65334.67	1400.47	13562.56	95085.19	
19	67072.80	1071.87	12171.67	93343.83	
4	73362.89	356.90	4562.66	92208.62	
26	83051.33	2819.15	5130.66	91174.16	
2	83884.74	623.16	5138.83	90236.78	
12	64597.87	165.91	3727.14	82917.19	
7	77428.46	616.92	11986.09	82575.84	
20	78000.04	508.22	10926.23	82667.16	
9	70616.81	421.14	18087.19	82409.55	
8	49450.79	1071.63	33799.24	80569.21	
3	58991.05	447.51	16042.76	78738.53	
21	77272.03	13.78	4473.01	78587.74	
11	68495.88	1045.28	19978.45	77726.83	
22	61195.49	338.16	28323.82	77368.68	
14	65294.08	255.85	6972.49	75197.59	
25	65418.72	41.28	7613.56	74120.48	
15	50526.27	185.80	20828.14	72421.99	
17	47355.38	309.99	17986.46	70624.15	
27	24463.02	99.14	40533.70	69880.01	
16	44589.54	302.40	19454.03	66173.78	
24	9633.45	864.17	70155.24	64635.68	
23	7464.79	1587.86	114779.45	65638.25	
5	19394.00	3549.02	12453.28	60454.39	
18	35049.85	38.94	10182.51	53295.53	
13	38145.62	31.16	2328.29	46731.78	
6	19442.10	342.59	43367.87	36710.68	
10	10880.20	1472.00	25250.98	32268.88	
28	5688.39	310.77	58379.77	27200.28	

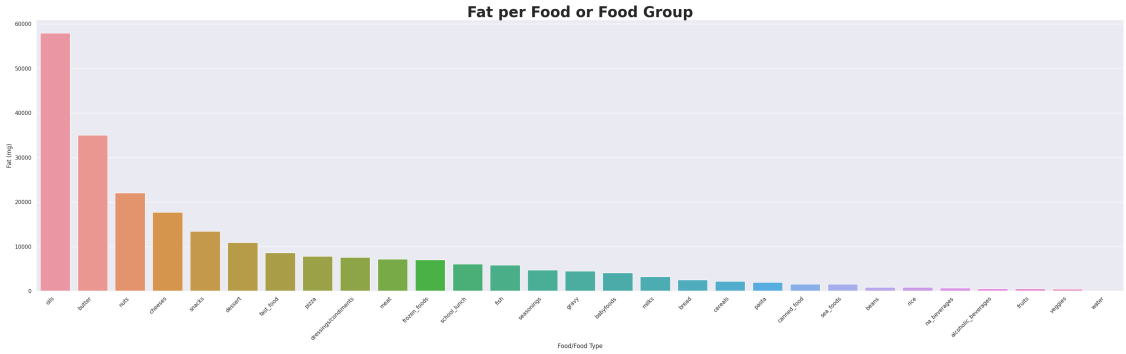
	Health_Score
0	99956.67
1	94799.29
19	92751.47
4	92127.09

```
26      90403.27
2       90178.93
12      82893.50
7       82344.08
20      82128.09
9       82080.77
8       79591.42
3       78466.88
21      78408.86
11      75809.78
22      75551.66
14      75131.39
25      73328.73
15      72123.67
17      70473.07
27      67047.31
16      65992.94
24      58672.05
23      54076.38
5       53516.31
18      52995.30
13      46623.73
6       35272.12
10      31538.79
28      24988.38
```

[4]:
```
#Create bar graph of Health Scores by Food/Food Group

nutrition_barplot = sns.barplot(x=all_averages['Name'], y =␣
 ↪all_averages['Health_Score'])
nutrition_barplot.set_ylabel('Health Score')
nutrition_barplot.set_xlabel('Food/Food Type')
nutrition_barplot.set_title('Health Score by Food/Food Group', fontdict={'size':
 ↪ 30, 'weight': 'bold'})
nutrition_barplot.set_xticklabels(nutrition_barplot.get_xticklabels(),␣
 ↪rotation=45, horizontalalignment='right')
plt.show()
```

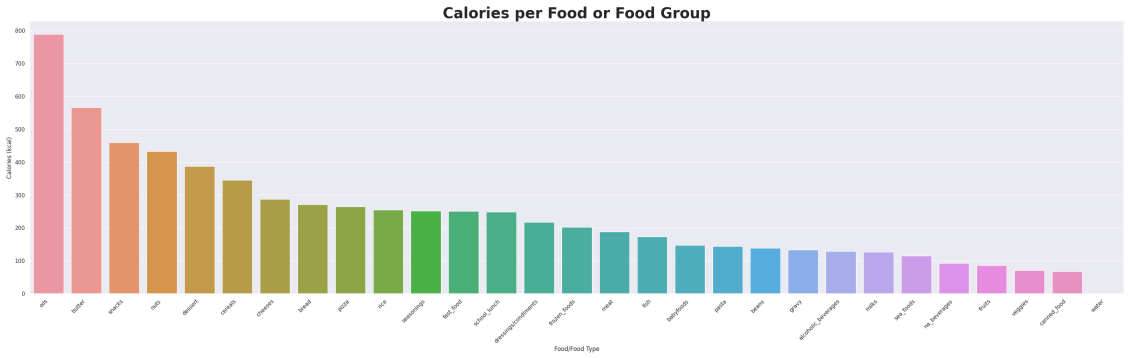


[5]:
```
#Create bar graph of Fat by Food/Food Group

all_averages = all_averages.sort_values(by='Total_Fat_(mg)', ascending=False)
nutrition_barplot = sns.barplot(x=all_averages['Name'], y =␣
 ↪all_averages['Total_Fat_(mg)'])
nutrition_barplot.set_ylabel('Fat (mg)')
nutrition_barplot.set_xlabel('Food/Food Type')
nutrition_barplot.set_title('Fat per Food or Food Group', fontdict={'size': 30,␣
 ↪'weight': 'bold'})
nutrition_barplot.set_xticklabels(nutrition_barplot.get_xticklabels(),␣
 ↪rotation=45, horizontalalignment='right')
plt.show()
```



[6]:
```
#Create bar graph of Calories by Food/Food Group

all_averages = all_averages.sort_values(by='Calories', ascending=False)
nutrition_barplot = sns.barplot(x=all_averages['Name'], y =␣
 ↪all_averages['Calories'])
nutrition_barplot.set_ylabel('Calories (kcal)')
nutrition_barplot.set_xlabel('Food/Food Type')
```

```
nutrition_barplot.set_title('Calories per Food or Food Group', fontdict={'size':
  ↪ 30, 'weight': 'bold'})
nutrition_barplot.set_xticklabels(nutrition_barplot.get_xticklabels(),␣
  ↪rotation=45, horizontalalignment='right')
plt.show()
```



0.2 Top 5 foods by Health Score

[7]:
```
df = df.sort_values(by='Health_Score', ascending=False)
top_health_score = df.loc[:, ['Name', 'Calories', 'Healthy_Nutrients',␣
  ↪'Unhealthy_Nutrients', 'Health_Score']]
top_health_score.head(10)
```

[7]:

	Name	Calories	Healthy_Nutrients	Unhealthy_Nutrients	Health_Score
1095	BEEF,NZ,IMP,VAR MEATS & BY-PRODUCTS,LIVER,RAW	133	216081.4	5489.0	216032.824779
1090	BEEF,NZ,IMP,VAR MEATS & BY-PRODUCTS LIVER,CKD,BLD	150	183358.4	6315.0	183284.105882
8530	VEAL,VAR MEATS&BY-PRODUCTS,LIVER,CKD,BRSD	192	183169.6	8757.0	183060.137500
4007	FISH OIL,COD LIVER	902	209282.2	123178.0	181909.311111
5161	LAMB,NZ,IMP,LIVER,CKD,SOAKED & FRIED	168	179570.3	9085.0	179463.417647
8531	VEAL,VAR MEATS&BY-PRODUCTS,LIVER,CKD,PAN-FRIED	193	177681.5	9104.0	177545.619403
5162	LAMB,NZ,IMP,LIVER,RAW	136	160920.4	6820.0	160860.046018
3760	DUCK,DOMESTICATED,LIVER,RAW	136	144492.5	6595.0	144342.613636

```
8532          144227.0              6744.0  143989.535211
8306          141448.4             11132.0  141238.362264
```

0.3 Bottom 5 foods by Health Score

[8]:
```
top_health_score.tail(10)
```

[8]:
```
                                                 Name  Calories  \
7955                                SUGAR,TURBINADO       399
7957                                   SUGARS,BROWN       380
7958                                SUGARS,GRANULATED       387
6225                    PEPPERS,SWT,GRN,FREEZE-DRIED       314
1688          BEVERAGES,COFFEE,INST,W/ WHTNR,RED CAL       509
8043    SWEETENERS,TABLETOP,SUCRALOSE,SPLENDA PACKETS       336
8042  SWEETENERS,TABLETOP,SACCHARIN (SODIUM SACCHARIN)       360
4413       HAZELNUTS,BEAKED (NORTHERN PLAINS INDIANS)       628
5858                    OIL,PAM COOKING SPRAY,ORIGINAL       792
8041                 SWEETENERS,TABLETOP,FRUCTOSE,LIQ       279

      Healthy_Nutrients  Unhealthy_Nutrients   Health_Score
7955               77.4             99190.0  -21485.643478
7957             1717.9             97020.0  -30622.100000
7958               24.0             99800.0  -35618.857143
6225            54879.4             41927.0  -49938.100000
1688            11980.6            113228.0  -54624.105882
8043             8834.0             80330.0  -71496.000000
8042            10122.0             85190.0  -75068.000000
4413            32455.5             52990.0 -100019.500000
5858            72366.0             83715.0 -206684.000000
8041            24003.2             76000.0 -735996.800000
```

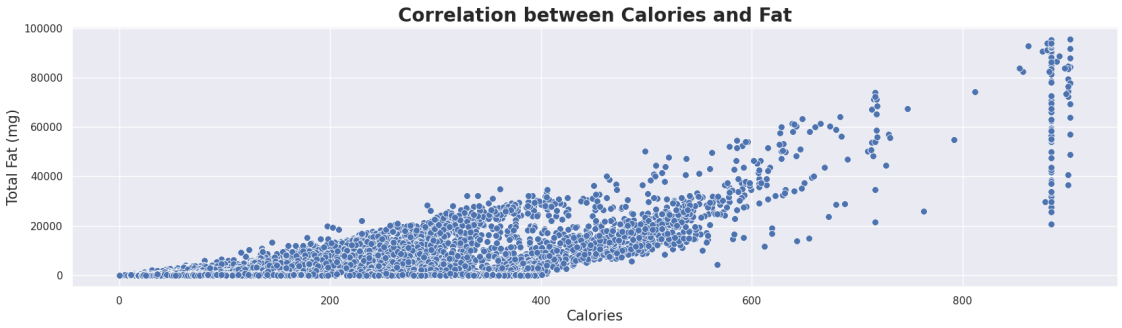
[9]:
```
#Create heatmap to check correlation in data
correlation = df[["Calories", 'Carbs_(mg)', 'Protein_(mg)', "Sugar_(mg)",␣
 ↪"Calcium_(mg)", "Total_Fat_(mg)", "Sodium_(mg)", "Potassium_(mg)",␣
 ↪"Water_(mg)", 'Vitamins_(mg)', 'Health_Score']].copy()
sns.set_theme(style="white")
corr = correlation.corr(method = 'pearson', min_periods = 1 )
corr.style.background_gradient(cmap='coolwarm')
```

[9]:
```
<pandas.io.formats.style.Styler at 0x7f881c4eef80>
```

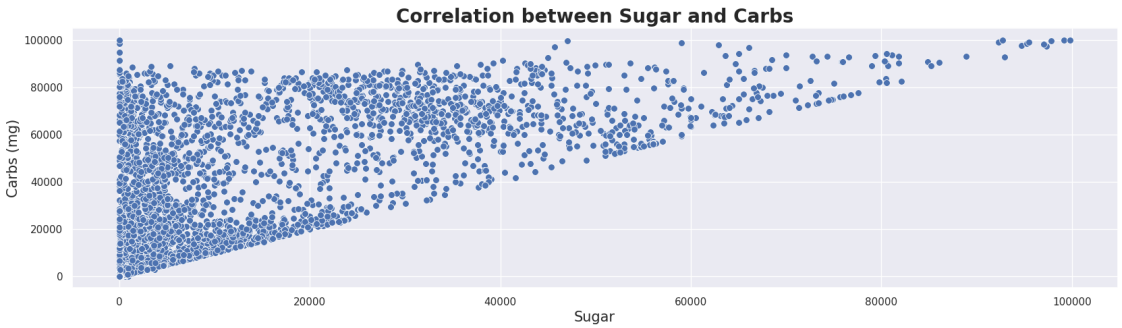
Notes: It appears that there isn't any strong correlations between any of the fields except Water's negative correlation with Calories. The highest positive correlation is between calories and fat, then sugar and carbs.

[10]:
```
#Scatter plot comparing Calories with Fat
sns.set(rc={"figure.figsize":(20, 5)})
```

```
scatter = sns.scatterplot(data=df, x='Calories', y='Total_Fat_(mg)',␣
 ↪legend='auto', s=50)
scatter.set_title("Correlation between Calories and Fat", fontdict={'size': 20,␣
 ↪'weight': 'bold'})
scatter.set_xlabel('Calories', fontdict={'size': 15})
scatter.set_ylabel('Total Fat (mg)', fontdict={'size': 15})
plt.show()
```

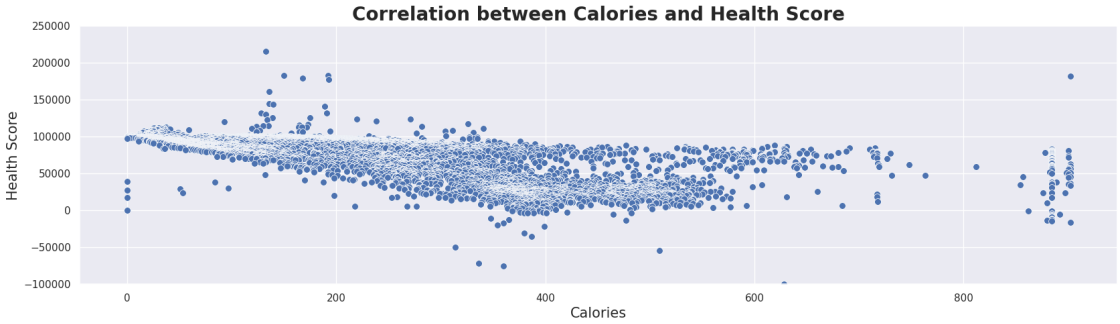


[11]:
```
#Scatter plot comparing sugar and carbs
sns.set(rc={"figure.figsize":(20, 5)})
scatter = sns.scatterplot(data=df, x='Sugar_(mg)', y='Carbs_(mg)',␣
 ↪legend='auto', s=50)
scatter.set_title("Correlation between Sugar and Carbs", fontdict={'size': 20,␣
 ↪'weight': 'bold'})
scatter.set_xlabel('Sugar', fontdict={'size': 15})
scatter.set_ylabel('Carbs (mg)', fontdict={'size': 15})
plt.show()
```



[12]:
```
#Scatter plot comparing Calories with Fat
sns.set(rc={"figure.figsize":(20, 5)})
scatter = sns.scatterplot(data=df, x='Calories', y='Health_Score',␣
 ↪legend='auto', s=50)
```

```
scatter.set_title("Correlation between Calories and Health Score",␣
 ↪fontdict={'size': 20, 'weight': 'bold'})
scatter.set_xlabel('Calories', fontdict={'size': 15})
scatter.set_ylabel('Health Score', fontdict={'size': 15})
plt.ylim(-100000, 250000)
plt.show()
```



0.4 Milk vs Cheese vs Butter

[13]:
```
#combine averages and transpose the data
mk_ch_btr_averages = pd.concat([milks_average, cheeses_average, butter_average])
mk_ch_btr_averages = mk_ch_btr_averages.transpose()
mk_ch_btr_averages.columns = ('Milk', 'Cheese', 'Butter')

#drop first two rows
mk_ch_btr_averages = mk_ch_btr_averages.iloc[2:]
mk_ch_btr_averages
```

[13]:

	Milk	Cheese	Butter
Calories	127.62	287.84	566.76
Protein_(mg)	8921.91	21502.7	15166.55
Carbs_(mg)	14776.17	5339.21	22978.97
Fiber_(mg)	70.21	134.48	4634.48
Sugar_(mg)	15879.44	1699.04	9829.55
Calcium_(mg)	300.55	581.78	130.31
Sodium_(mg)	175.49	754.25	277.41
Total_Fat_(mg)	3290.34	17747.44	35011.24
Potassium_(mg)	406.04	145.93	441.07
Water_(mg)	70616.81	49450.79	9633.45
Vitamins_(mg)	421.14	1071.63	864.17
Unhealthy_Nutrients	18087.19	33799.24	70155.24
Healthy_Nutrients	82409.55	80569.21	64635.68
Health_Score	82080.77	79591.42	58672.05

Notes: Milk has the highest health score, followed by Cheese in a close second and Butter in a distant third. Butter has by far the highest calories.

0.5 Analyze Milk: Are "lowfat milks" really low in fat?

[14]:
```
#Combine the two and compare
compare_milk = pd.concat([lowfat_milk.loc['9998'], other_milk.loc['9999']],␣
 ↪axis=1)
compare_milk.rename(columns={'9998': 'Low_Fat_Milk_Average', '9999':␣
 ↪'Other_Milk_Average'}, inplace=True)

#drop first two rows
compare_milk = compare_milk.iloc[2:]
compare_milk
```

[14]:

	Low_Fat_Milk_Average	Other_Milk_Average
Calories	137.88	121.8
Protein_(mg)	13220.0	6486.33
Carbs_(mg)	19556.47	12067.33
Fiber_(mg)	5.88	106.67
Sugar_(mg)	21639.17	12999.58
Calcium_(mg)	414.65	235.9
Sodium_(mg)	310.53	98.97
Total_Fat_(mg)	584.12	4823.87
Potassium_(mg)	578.71	308.2
Water_(mg)	63645.88	74567.0
Vitamins_(mg)	612.37	312.78
Unhealthy_Nutrients	16355.24	19068.63
Healthy_Nutrients	79391.2	84119.95
Health_Score	79055.12	83795.31

Notes: The Low Fat Milk option does have less fat than other milks, but it also has a lower health score and more calories.

0.6 Butter vs Margarine

[15]:
```
#combine averages and transpose the data
butter_margarine_averages = pd.concat([butter_average, margarine_average])
butter_margarine_averages = butter_margarine_averages.transpose()
butter_margarine_averages.columns = ('Butter', 'Margarine')

#drop first two rows
butter_margarine_averages = butter_margarine_averages.iloc[2:]
butter_margarine_averages
```

[15]:

	Butter	Margarine
Calories	566.76	702.5

Protein_(mg)	15166.55	370.0
Carbs_(mg)	22978.97	776.43
Fiber_(mg)	4634.48	0.0
Sugar_(mg)	9829.55	0.0
Calcium_(mg)	130.31	10.83
Sodium_(mg)	277.41	582.86
Total_Fat_(mg)	35011.24	53190.5
Potassium_(mg)	441.07	28.42
Water_(mg)	9633.45	18163.57
Vitamins_(mg)	864.17	3758.8
Unhealthy_Nutrients	70155.24	94116.93
Healthy_Nutrients	64635.68	83658.66
Health_Score	58672.05	75975.19

Notes: Margarine has a MUCH higher health score, but also more calories compared to Butter. Butter also has less fat and much more protein.

0.7 Compare Fruits and Veggies

[16]:
```
#combine averages and transpose the data
fruits_veggies_averages = pd.concat([fruits_average, veggies_average])
fruits_veggies_averages = fruits_veggies_averages.transpose()
fruits_veggies_averages.columns = ('Fruits', 'Veggie')

#drop first two rows
fruits_veggies_averages = fruits_veggies_averages.iloc[2:]
fruits_veggies_averages
```

[16]:

	Fruits	Veggie
Calories	85.8	70.75
Protein_(mg)	1535.27	2258.19
Carbs_(mg)	19218.96	10116.72
Fiber_(mg)	2201.82	2330.41
Sugar_(mg)	13534.41	5298.86
Calcium_(mg)	34.25	59.56
Sodium_(mg)	33.19	138.52
Total_Fat_(mg)	581.96	474.95
Potassium_(mg)	233.8	286.5
Water_(mg)	77428.46	83051.33
Vitamins_(mg)	616.92	2819.15
Unhealthy_Nutrients	11986.09	5130.66
Healthy_Nutrients	82575.84	91174.16
Health_Score	82344.08	90403.27

Notes: Veggies tend to be lower in calories, carbs, sugar, and fat, while much higher in Sodium and Vitamins compared to Fruit. Fruit has a slightly higher Health Score.

0.8 Compare Meats

[25]:
```
meat_averages = meat_averages.drop('id', axis=1)
meat_averages
```

[25]:

	Name	Calories	Protein_(mg)	Carbs_(mg)	Fiber_(mg)
0	cow_Average	208.22	23583.69	639.42	48.18
1	turkeys_average	175.18	19960.61	1855.49	122.98
2	pig_Average	217.41	20303.11	1916.52	121.98
3	chicken_Average	223.48	21617.82	3167.34	186.70
4	game_Average	157.88	26626.79	0.00	0.00
5	lamb_Average	223.54	22374.97	129.04	9.02
6	veal_Average	185.06	24004.19	381.24	5.15
7	emu_Average	133.00	26106.67	0.00	0.00
8	ostrich_Average	135.94	24553.33	0.00	0.00
9	duck_Average	208.77	19182.31	5117.69	66.67
10	goose_Average	218.00	21041.43	1095.71	0.00
11	fish_Average	173.93	18668.08	3459.69	335.14
9999	meat_average	188.37	22335.25	1480.18	74.65

	Sugar_(mg)	Calcium_(mg)	Sodium_(mg)	Total_Fat_(mg)	Potassium_(mg)
0	93.14	13.77	104.98	9973.41	309.23
1	366.97	25.77	337.95	6224.00	227.52
2	498.10	18.90	502.15	10824.33	309.38
3	345.36	25.90	244.57	9059.23	221.46
4	0.00	10.43	61.56	3167.92	353.57
5	14.46	13.93	70.87	11542.55	270.58
6	21.92	16.30	94.10	7089.15	297.76
7	0.00	5.33	97.67	1600.42	321.67
8	0.00	5.72	77.22	2378.17	332.78
9	5245.00	13.54	117.08	8870.08	221.33
10	313.33	22.71	90.57	10196.86	317.29
11	485.31	43.57	330.19	5847.88	345.68
9999	615.30	17.99	177.41	7231.17	294.02

	Water_(mg)	Vitamins_(mg)	Unhealthy_Nutrients	Healthy_Nutrients
0	62967.46	362.39	16840.22	93356.28
1	67553.66	941.49	12632.34	95181.03
2	62104.49	319.04	19163.94	91662.51
3	60937.07	912.00	17444.12	92633.69
4	67109.46	1.60	6489.80	96372.87
5	62570.21	890.95	20720.43	92702.21
6	65380.86	2211.90	12828.42	96551.20
7	70396.67	6.61	3129.08	98572.05
8	72170.00	0.94	4693.11	99163.60
9	62877.69	4150.74	18145.92	93382.90
10	62875.71	5936.09	18491.71	98100.07

```
11      67072.80        1071.87             12171.67          93343.83
9999    65334.67        1400.47             13562.56          95085.19
```

```
      Health_Score
0          93092.65
1          94801.80
2          91192.69
3          92149.00
4          96263.77
5          92329.28
6          96303.39
7          98537.90
8          99110.47
9          93131.78
10         97927.28
11         92751.47
9999       94799.29
```

Notes: Ostrich and Goose have the highest Health Scores, Chicken contains the most calories, Game and Emu contain the most protein, Lamb contains the most fat, Pig contains (by far) the most sodium, and fish contains the most vitamins and potassium.

0.9 Compare Alcoholic Beverages, Non-Alcohol Beverages, and Water

[18]:
```
#combine averages and transpose the data
beverage_averages = pd.concat([na_beverages_average,␣
 ↪alcoholic_beverages_average, water_average])
beverage_averages = beverage_averages.transpose()
beverage_averages.columns = ('Non-Alcoholic Beverages', 'Alcoholic Beverages',␣
 ↪'Water')

#drop first two rows
beverage_averages = beverage_averages.iloc[2:]
beverage_averages
```

[18]:

	Non-Alcoholic Beverages	Alcoholic Beverages	Water
Calories	92.27	129.8	0.33
Protein_(mg)	2769.25	1025.27	0.0
Carbs_(mg)	15503.54	7737.3	43.33
Fiber_(mg)	618.89	127.08	0.0
Sugar_(mg)	11182.85	7724.59	0.0
Calcium_(mg)	57.69	12.33	5.0
Sodium_(mg)	98.32	25.35	5.67
Total_Fat_(mg)	750.12	598.46	0.0
Potassium_(mg)	233.84	82.46	1.33
Water_(mg)	78000.04	77272.03	99943.33

Vitamins_(mg)	508.22	13.78	0.0
Unhealthy_Nutrients	10926.23	4473.01	0.0
Healthy_Nutrients	82667.16	78587.74	99956.67
Health_Score	82128.09	78408.86	99956.67

Notes: Non-Alcoholic Beverages have a higher Health Score, but generally have more carbs and sugar, while Alcoholic Beverages have more calories and sodium.

0.10 Compare Baby Food, fast Food, Frozen Foods, Canned Foods, and School Lunches

[19]:
```
#combine averages and transpose the data
bf_ff_sl_averages = pd.concat([babyfoods_average, fast_food_average,␣
 ↪frozen_food_average, canned_food_average, school_lunches_average])
bf_ff_sl_averages = bf_ff_sl_averages.transpose()
bf_ff_sl_averages.columns = ('Baby Foods', 'Fast Foods', 'Frozen Foods',␣
 ↪'Canned Food', 'School Lunches')

#drop first two rows
bf_ff_sl_averages = bf_ff_sl_averages.iloc[2:]
bf_ff_sl_averages
```

[19]:

	Baby Foods	Fast Foods	Frozen Foods	Canned Food	School Lunches
Calories	147.75	251.4	202.28	67.94	249.15
Protein_(mg)	4015.73	12195.61	10628.27	2865.58	13523.08
Carbs_(mg)	21141.52	22419.04	18796.29	9277.73	27435.38
Fiber_(mg)	894.1	1435.52	2448.46	1315.06	3430.77
Sugar_(mg)	13834.81	4849.23	4111.21	2495.42	4965.38
Calcium_(mg)	135.63	98.09	78.77	17.5	165.0
Sodium_(mg)	54.27	549.37	370.55	455.42	461.46
Total_Fat_(mg)	4103.28	8671.48	7029.89	1617.01	6113.62
Potassium_(mg)	192.32	211.08	233.85	231.81	
Water_(mg)	68495.88	50526.27	58991.05	83884.74	
Vitamins_(mg)	1045.28	185.8	447.51	623.16	
Unhealthy_Nutrients	19978.45	20828.14	16042.76	5138.83	
Healthy_Nutrients	77726.83	72421.99	78738.53	90236.78	
Health_Score	75809.78	72123.67	78466.88	90178.93	

Potassium_(mg)	309.31
Water_(mg)	47355.38
Vitamins_(mg)	309.99
Unhealthy_Nutrients	17986.46
Healthy_Nutrients	70624.15
Health_Score	70473.07

Notes: Canned foods have by far the highest health score, followed by School Lunches. Nutrients in Fast food and School Lunches are comparable on all metrics but calcium, while baby foods are lowest in most metrics, but very high in vitamins

0.11 Chinese Food vs Sea Food vs Pasta vs Pizza

[20]:
```
#combine averages and transpose the data
cf_sf_pas_pz_averages = pd.concat([chinese_foods_average, seafoods_average,␣
 ↪pasta_average, pizza_average])
cf_sf_pas_pz_averages = cf_sf_pas_pz_averages.transpose()
cf_sf_pas_pz_averages.columns = ('Chinese Food', 'Sea Food', 'Pasta', 'Pizza')

#drop first two rows
cf_sf_pas_pz_averages = cf_sf_pas_pz_averages.iloc[2:]
cf_sf_pas_pz_averages
```

[20]:

	Chinese Food	Sea Food	Pasta	Pizza
Calories	146.45	115.0	144.96	265.2
Protein_(mg)	4950.79	14444.21	5615.76	12013.45
Carbs_(mg)	20926.05	6506.58	24432.64	29713.1
Fiber_(mg)	1367.65	940.58	2256.03	2668.97
Sugar_(mg)	3820.0	673.11	3796.94	4031.22
Calcium_(mg)	30.74	54.7	41.87	178.46
Sodium_(mg)	257.0	486.53	305.26	584.67
Total_Fat_(mg)	2189.65	1573.08	1971.51	7849.79
Potassium_(mg)	196.18	282.22	197.8	212.55
Water_(mg)	68062.37	73362.89	65294.08	44589.54
Vitamins_(mg)	704.21	356.9	255.85	302.4
Unhealthy_Nutrients	8594.21	4562.66	6972.49	19454.03
Healthy_Nutrients	79025.78	92208.62	75197.59	66173.78
Health_Score	78956.59	92127.09	75131.39	65992.94

Notes: Sea Food has easily the highest health score and lowest calories. Chinese Food and Pasta are close for second place, while Pizza is in a distant last.

0.12 Desserts vs Snacks vs Cereal

[21]:
```
#combine averages and transpose the data
des_snck_averages = pd.concat([dessert_average, snacks_average,␣
 ↪cereals_average])
des_snck_averages = des_snck_averages.transpose()
des_snck_averages.columns = ('Desserts', 'Snacks', 'Cereal')

#drop first two rows
des_snck_averages = des_snck_averages.iloc[2:]
des_snck_averages
```

[21]:

	Desserts	Snacks	Cereal
Calories	387.99	459.4	345.15
Protein_(mg)	5355.54	6928.56	8340.4
Carbs_(mg)	57653.35	65323.22	74043.8
Fiber_(mg)	2654.16	3893.62	7722.29
Sugar_(mg)	25868.63	34553.79	20624.17
Calcium_(mg)	97.09	132.37	159.3
Sodium_(mg)	374.59	299.48	382.33
Total_Fat_(mg)	10904.0	13444.63	2243.61
Potassium_(mg)	200.54	335.94	274.86
Water_(mg)	19442.1	5688.39	10880.2
Vitamins_(mg)	342.59	310.77	1472.0
Unhealthy_Nutrients	43367.87	58379.77	25250.98
Healthy_Nutrients	36710.68	27200.28	32268.88
Health_Score	35272.12	24988.38	31538.79

Notes: Desserts have the highest health score, while cereal has the lowest calories.

0.13 Dressings vs Oils vs Seasonings vs Gravy

[22]:
```
#combine averages and transpose the data
drs_oil_ssngs_averages = pd.concat([dressings_average, oils_average,␣
 ↪seasonings_average, gravy_average])
drs_oil_ssngs_averages = drs_oil_ssngs_averages.transpose()
drs_oil_ssngs_averages.columns = ('Dressings', 'Oils', 'Seasonings', 'Gravy')

#drop first two rows
drs_oil_ssngs_averages = drs_oil_ssngs_averages.iloc[2:]
drs_oil_ssngs_averages
```

[22]:

	Dressings	Oils	Seasonings	Gravy
Calories	217.7	789.23	251.84	134.15
Protein_(mg)	2334.84	1993.65	8406.4	4158.97
Carbs_(mg)	17902.79	3052.29	50282.0	20772.56
Fiber_(mg)	1025.41	164.89	21264.0	1148.0

Sugar_(mg)	10872.0	189.23	3622.63	4053.18
Calcium_(mg)	41.66	9.59	597.52	62.21
Sodium_(mg)	895.73	120.05	2752.4	1903.54
Total_Fat_(mg)	7594.59	57971.22	4744.64	4530.85
Potassium_(mg)	152.72	72.05	1071.35	207.4
Water_(mg)	61195.49	7464.79	19394.0	65418.72
Vitamins_(mg)	338.16	1587.86	3549.02	41.28
Unhealthy_Nutrients	28323.82	114779.45	12453.28	7613.56
Healthy_Nutrients	77368.68	65638.25	60454.39	74120.48
Health_Score	75551.66	54076.38	53516.31	73328.73

Notes: Dressings have the highest health score, Gravy has the lowest calories, Oils has by far the highest calories.

0.14 Nuts vs Bread vs Beans

[23]:
```
#combine averages and transpose the data
nts_brd_bns_averages = pd.concat([nuts_average, bread_average, beans_average])
nts_brd_bns_averages = nts_brd_bns_averages.transpose()
nts_brd_bns_averages.columns = ('Nuts', 'Bread', 'Beans')

#drop first two rows
nts_brd_bns_averages = nts_brd_bns_averages.iloc[2:]
nts_brd_bns_averages
```

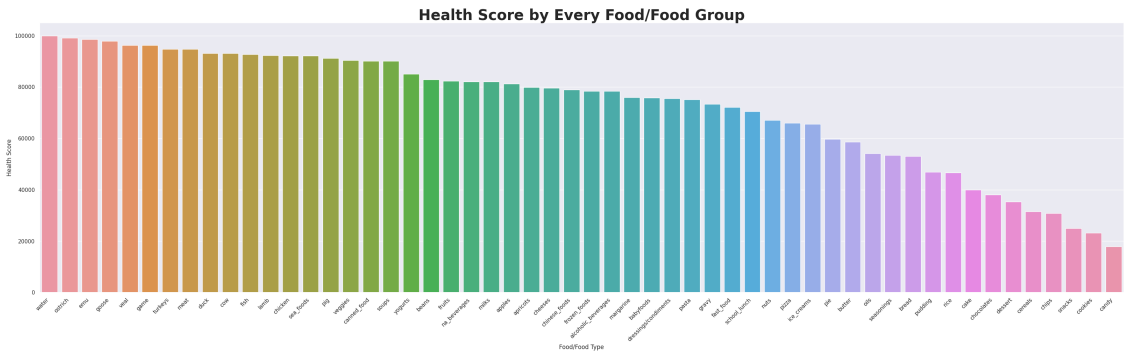
[23]:

	Nuts	Bread	Beans
Calories	433.19	271.84	139.41
Protein_(mg)	12127.14	9474.18	9366.28
Carbs_(mg)	28429.21	48594.48	22463.09
Fiber_(mg)	7612.24	5084.13	7245.93
Sugar_(mg)	6714.17	5071.9	2188.75
Calcium_(mg)	81.38	89.91	62.24
Sodium_(mg)	107.7	460.52	139.57
Total_Fat_(mg)	22034.63	2531.25	900.38
Potassium_(mg)	581.07	188.97	477.95
Water_(mg)	24463.02	35049.85	64597.87
Vitamins_(mg)	99.14	38.94	165.91
Unhealthy_Nutrients	40533.7	10182.51	3727.14
Healthy_Nutrients	69880.01	53295.53	82917.19
Health_Score	67047.31	52995.3	82893.5

Notes: Beans have by far the highest health score, followed by Nuts in second and Bread in a distant third. Nuts have by far the most calories, but also the most protein. Beans seem to be high in vitamins.

0.15 Compare EVERYTHING

[24]:
```
averages_list = [milks_average, water_average, canned_food_average,␣
 ↪fruits_average, cheeses_average, soups_average, cereals_average,␣
 ↪cookies_average, babyfoods_average, beans_average, rice_average,␣
 ↪ice_creams_average, chips_average, pasta_average, fast_food_average,␣
 ↪school_lunches_average, bread_average, fish_average, candy_average,␣
 ↪na_beverages_average, alcoholic_beverages_average, dressings_average,␣
 ↪yogurts_average, oils_average, cake_average, butter_average,␣
 ↪chicken_average, duck_average, goose_average, ostrich_average, emu_average,␣
 ↪cow_average, pig_average, lamb_average, veal_average, game_average,␣
 ↪gravy_average, pie_average, pudding_average, soups_average, apples_average,␣
 ↪apricots_average, veggies_average, nuts_average, chocolates_average,␣
 ↪snacks_average, dessert_average, seasonings_average, chinese_foods_average,␣
 ↪seafoods_average, frozen_food_average, turkey_average, meat_average,␣
 ↪pizza_average, margarine_average,]

#combine all the dataframes into one using concat
all_averages = pd.concat(averages_list, ignore_index=True)

all_averages = all_averages.sort_values(by='Health_Score', ascending=False)
all_averages['Name'] = all_averages['Name'].str.replace('_average', '')
all_averages['Name'] = all_averages['Name'].str.replace('_Average', '')

#Create bar graph of top ten Countries by coffee consumption
sns.set(rc={"figure.figsize":(40, 10)})
nutrition_barplot = sns.barplot(x=all_averages['Name'], y =␣
 ↪all_averages['Health_Score'])
nutrition_barplot.set_ylabel('Health Score')
nutrition_barplot.set_xlabel('Food/Food Type')
nutrition_barplot.set_title('Health Score by Every Food/Food Group',␣
 ↪fontdict={'size': 30, 'weight': 'bold'})
nutrition_barplot.set_xticklabels(nutrition_barplot.get_xticklabels(),␣
 ↪rotation=45, horizontalalignment='right')
#plt.ylim(0,50000)
plt.show()
```



Notes: Unsurprising results, with health score near the top being water, live animal products high in protein, then vegetables and fruits. Toward the middle and lower parts, you have manufactured products like beverages, dressings/condiments, frozen and baby foods, pasta, and alcohol. You also have animal byproducts like milk, cheese, and margarine. Leading towards the bottom of health scores, you have fast food and school lunches, pizza, oils, seasonings, rice, sweets. and various snacks.