McDonald's Competition

ABOUT THE COMPANY:

McDonald's Corporation is an American fast food company founded 80 years ago in California, U.S and is the second-largest private employer. It is mainly known for its variety of burgers and fries but due to wake of "junk food concept", it has added salads, smoothies etc., to is menu. Though it started in the U.S., it has expanded itself widely all over the world with more than 38,000 locations due to the company's better understanding of consumer's taste & preference and market. It's revenue (as of 2019) stood at $21.2 billion. The company's main competitors are: Burger King, Wendy's and KFC.

ABOUT THE DATA:

The data contains 24 columns which are:

1. Category: The menu is broadly divided as Breakfast, Beef or Pork, Chicken or Fish, Smoothies etc.
2. Item: Under paricular category, they show us the sub-items like EggMcMuffin under Breakfast, Big Mac under Beef or Pork etc.,
3. Serving Size: What is the serving size of each item is shown here
4. Calories: Calories derived from each item
5. Calories from fat
6. Total Fat: Fat contained in each item
7. Total Fat (% Daily Value)
8. Saturated Fat: It is one of the unhealthy fats, this shows us the saturated fat contained in the respective items
9. Saturated Fat (% Daily Value)
10. Trans Fat: It is also considered one of the unhealthy fats and shows us the numbers for items given
11. Cholesterol: A greasy substance; Generally lower it is, better it is the item.
12. Cholesterol (% Daily Value)
13. Sodium: Derived from salt, sauces etc., higher intake may have severe consequences on health
14. Sodium (% Daily Value)
15. Carbohydrates: an important source of energy and is found in both healthy and unhealthy foods; this shows the carbohydreates content in a given item
16. Carbohydrates (% Daily Value)
17. Dietary Fiber: shows the dietary fiber content (dietary fiber is also known as roughage)
18. Dietary Fiber (% Daily Value)
19. Sugars
20. Protein
21. Vitamin A (% Daily Value)
22. Vitamin C (% Daily Value)
23. Calcium (% Daily Value)
24. Iron (% Daily Value)

In [1]:
```
#IMPORTING NECESSARY LIBRARIES
import pandas as pd
import numpy as np
import os
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(color_codes=True)
%matplotlib inline
```

In [2]:
```
#Reading the 'csv' file
data=pd.read_csv('Mcdonald .csv')
```

Plot graphically which food categories have the highest and lowest varieties

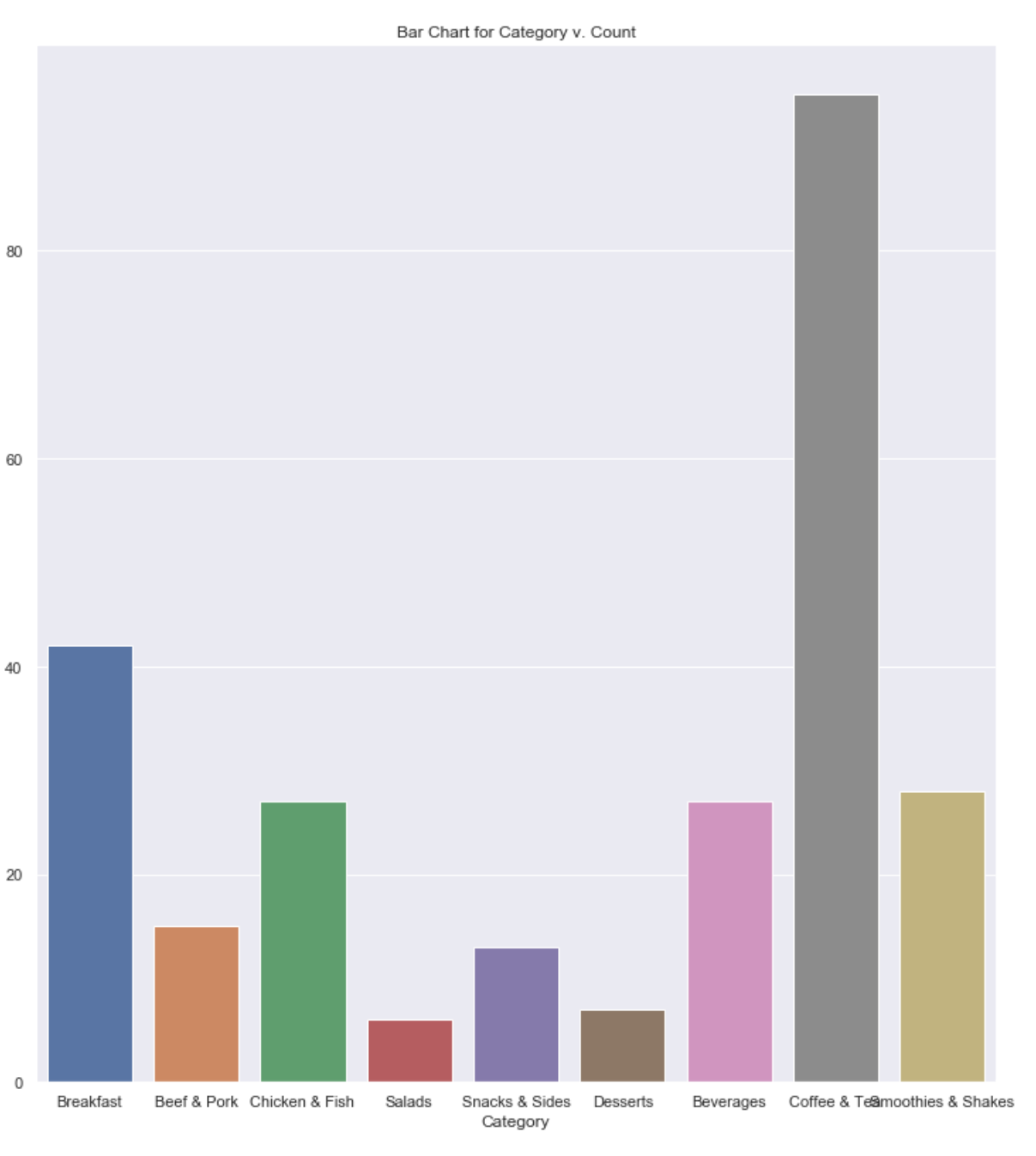
In [3]:
```
#'value_counts() will sum up the values in the 'Category'
data_cat=data['Category'].value_counts()
data_cat
#as we can see the highest variety is in 'Coffee & Tea' and lowest in
#'Salads'
```

Out[3]:
```
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 [4]:
```
#the above table can also be graphically represented as shown:
plt.figure(figsize=(14,16))
plot_1=sns.catplot(x="Category", kind="count", data=data)
plot_1.fig.set_figwidth(12)
plot_1.fig.set_figheight(12)
plt.title("Bar Chart for Category v. Count")
plt.xlabel("Category")
plt.ylabel("Count")
#Coffee & Tea has the highest variety of items contained in it- it sells
500 million cups per day in U.S alone and hence justifies the number of
 varieties available under this^
#Salads has the lowest variety of items contained in it- this may be due
to the fact that it will be less preferred by people and only contribute
2-3% of sales*
#^https://www.factretriever.com/mcdonalds-food-facts
#*https://www.huffingtonpost.in/entry/mcdonalds-salads-sales_n_3355068?r
i18n=true
```

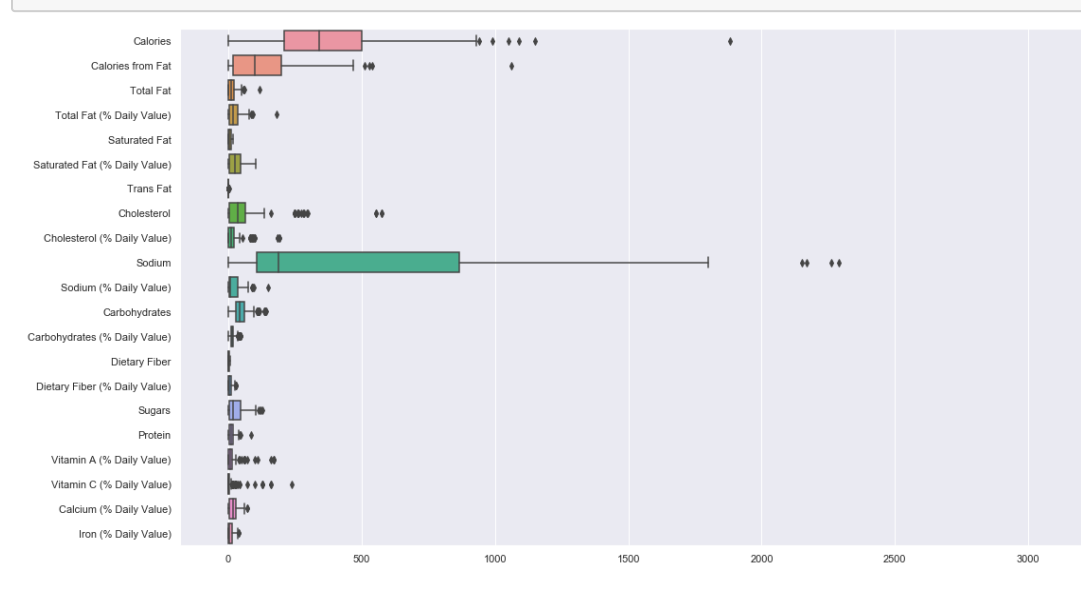
Out[4]:
```
Text(9.924999999999997, 0.5, 'Count')
```

```
<Figure size 1008x1152 with 0 Axes>
```



Which all variables have an outlier?

In [5]:
```
#Outlier- which significantly differs from other observations in a given
variable
fig, ax = plt.subplots(figsize=(20,10))
sns.boxplot(data=data,orient="h")
plt.show()
#All the variables except 'Saturated Fat' & 'Dietary Fiber' have
#outliers
```



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

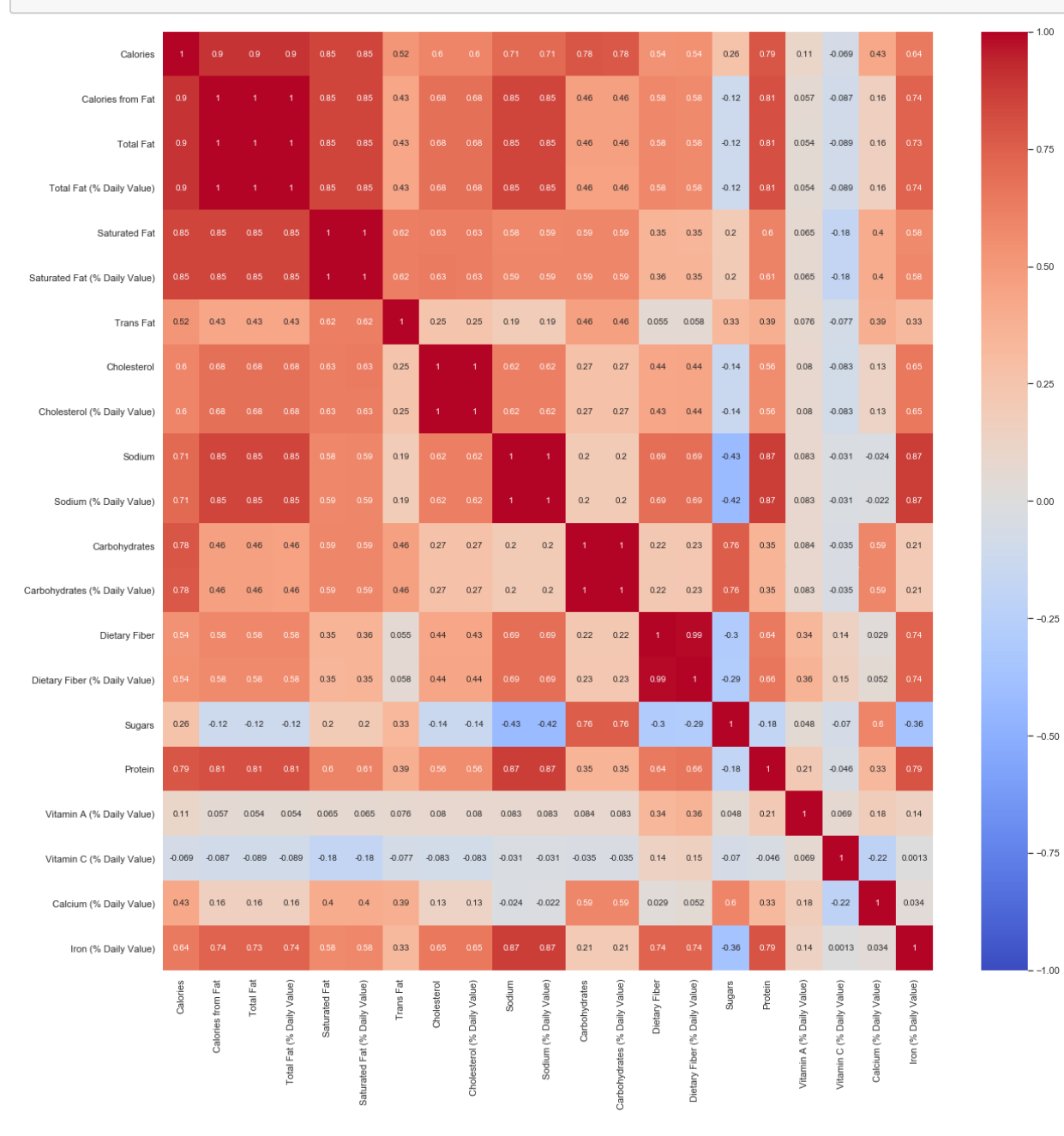
In [6]:
```
#To find the correlation matrix
cor = data.corr()
cor
```

Out[6]:

	Calories	Calories from Fat	Total Fat	Total Fat (% Daily Value)	Saturated Fat	Saturated Fat (% Daily Value)	Trans Fat	Cholest
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
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 [7]:
```
plt.figure(figsize=(20,20))
sns.heatmap(cor,annot=True, cmap='coolwarm',vmin=-1,vmax=1)
plt.show()
#Ignoring the variables which are almost the same such as Total Fat &
#Total Fat (% Daily Value), Cholesterol & Cholesterol (% Daily Value)
#and the diagonal values (which will always be 1), we see that there
#is a high correlation between Total Fat & Calories.
#we see that there is a high correlation between Total Fat & Calories
#Considering Total Fat and Calories from Fat (whose correlation=0.999)
#will according to me won't be justifiable as: though Calories & Fat
#are two different concepts, but 'Calories from Fat' is very closely
#related to 'Total Fat' as well as 'Saturated Fat' (but the
#correlation between Saturated Fat and Calories from Fat is relatively
#lesser than that of Total Fat)
```



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

In [8]:
```
data_df = data[['Cholesterol (% Daily Value)', 'Category']].groupby('Cat
egory').sum()
data_df
```

Out[8]:

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

In [9]:
```
total = data_df.sum(numeric_only=True)[0]
data_df['percentage'] = data_df["Cholesterol (% Daily Value)"]/ total *1
00
print (data_df)
```

```
                    Cholesterol (% Daily Value)  percentage
Category
Beef & Pork                                 434    9.075701
Beverages                                     5    0.104559
Breakfast                                  2140   44.751150
Chicken & Fish                              681   14.240903
Coffee & Tea                                891   18.632371
Desserts                                     34    0.711000
Salads                                      104    2.174822
Smoothies & Shakes                          412    8.615642
Snacks & Sides                               81    1.693852
```

In [10]:
```
df_3 = pd.DataFrame(data_df.groupby(['Category', 'percentage']).size()).
sort_values(by='percentage',ascending=False).head(1)
df_3
#Under the 'Breakfast' category, the items contained are: Bacon, Egg &
#Cheese Biscuit with Egg Whites (Regular Biscuit), Sausage Biscuit
#with Egg Whites (Regular Biscuit) etc.,
#as we know, Bacon which itself contains approximately 35% cholesterol
#per 100 gms (According to UDSA), Egg contains 124% cholesterol per 100
 gms
#(According to USDA), Cheese & Sausages with approx 30% cholesterol per
 100 gms
#this may have contributed to the increase of Cholesterol percentage in
 the
#above table
#on the other hand, the cholesterol levels of beef, pork, chicken etc
#contains less than 30% cholesterol per 100 gms each, which explains
#the percentage we see above.
```

Out[10]:

		0
Category	**percentage**	
Breakfast	**44.75115**	1

Which item contributes maximum to the Sodium intake?

In [11]:
```
data_df_1 = data[['Sodium', 'Item']].groupby('Item').sum()
data_df_1.sort_values(by="Sodium",ascending=False).head(1)
# A 10-piece order of Chicken McNuggets packs in 900 milligrams (mg) of
#sodium, more than half the sodium you should have in a single day
#approx 151% of sodium is contained in 'Chicken McNuggets (40 piece)'
#as compared to 'Big Breakfast with Hotcakes and Egg Whites (Large Biscu
it)'
#which has the next highest amount of sodium content and is ranked as
#the 9th healthiest breakfast item in McD's menu
#Big Breakfast with Hotcakes (Large Biscuit) which stands third in the
#amount of sodium content is interestingly ranked as the 2nd healthiest
#breakfast item.
#Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit) which
#comes under top 4 amount of sodium content and apparently ranked as
#the 3rd healthiest breakfast item.**
#**https://www.mercurynews.com/2016/05/13/mcdonalds-breakfast-items-rank
ed-by-healthiness/
```

Out[11]:

	Sodium
Item	
Chicken McNuggets (40 piece)	3600

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

In [12]:
```
data_df_2 = data[['Saturated Fat', 'Item']].groupby('Item').sum()
data_df_2.sort_values(by="Saturated Fat",ascending=False).head(4)
#the food which are high in saturated fat are milk and white chocolate,
#toffee, cakes, puddings and biscuits, pastries and pies fatty meat,
#butter, lard, ghee*
#this explains why the top items under 'Saturated Fat' are mostly 'oily'
#or 'sweet' items (under broad classification-Smoothies, Coffee & Tea)
#*https://www.heartuk.org.uk/low-cholesterol-foods/saturated-fat
```

Out[12]:

	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