McDonald's Nutritional Analysis

Shweta Kumari https://github.com/shweta021119/McDonald-s-Nutritional-Analysis.git (https://github.com/shweta021119/McDonald-s-Nutritional-Analysis.git)

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
import os
```

In [3]:
```
os.getcwd()
```

Out[3]:
```
'C:\\Users\\Lenovo\\Untitled Folder 2'
```

In [4]:
```
data = r"E:\bootcamp"
```

In [5]:
```
os.chdir(data)
```

In [6]:
```
os.getcwd()
```

Out[6]:
```
'E:\\bootcamp'
```

In [7]:
```
df = pd.read_csv("Nutrical Dataset.csv")
```

In [175]:
```
df.head(10)
```

Out[175]:

urated Fat	Saturated Fat (% Daily Value)	Trans Fat	...	Carbohydrates	Carbohydrates (% Daily Value)	Dietary Fiber	Dietary Fiber (% Daily Value)	Sugars	Protein	Vitamin A (% Daily Value)	Vitan C Da Valı
5.0	25	0.0	...	31	10	4	17	3	17	10	
3.0	15	0.0	...	30	10	4	17	3	18	6	
8.0	42	0.0	...	29	10	4	17	2	14	8	
10.0	52	0.0	...	30	10	4	17	2	21	15	
8.0	42	0.0	...	30	10	4	17	2	21	6	
9.0	46	1.0	...	31	10	4	18	3	26	15	
13.0	65	0.0	...	38	13	2	7	3	19	10	
14.0	68	0.0	...	43	14	3	12	4	19	15	
11.0	56	0.0	...	36	12	2	7	3	20	2	
12.0	59	0.0	...	42	14	3	12	4	20	6	

In [9]:
```
df.shape
```

Out[9]: (260, 24)

In [10]:
```
df.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: 48.9+ KB
```

In [15]:
```
df.columns
```

Out[15]:
```
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 [13]:
```
df.isnull().sum()
```

Out[13]:
```
Category                          0
Item                              0
Serving Size                      0
Calories                          0
Calories from Fat                 0
Total Fat                         0
Total Fat (% Daily Value)         0
Saturated Fat                     0
Saturated Fat (% Daily Value)     0
Trans Fat                         0
Cholesterol                       0
Cholesterol (% Daily Value)       0
Sodium                            0
Sodium (% Daily Value)            0
Carbohydrates                     0
Carbohydrates (% Daily Value)     0
Dietary Fiber                     0
Dietary Fiber (% Daily Value)     0
Sugars                            0
Protein                           0
Vitamin A (% Daily Value)         0
Vitamin C (% Daily Value)         0
Calcium (% Daily Value)           0
Iron (% Daily Value)              0
dtype: int64
```

In [16]:
```
df.duplicated()
```

Out[16]:
```
0      False
1      False
2      False
3      False
4      False
       ...
255    False
256    False
257    False
258    False
259    False
Length: 260, dtype: bool
```

In [17]:
```
df.duplicated().sum()
```

Out[17]:
```
0
```

3. Exploratory Data Analysis (EDA):

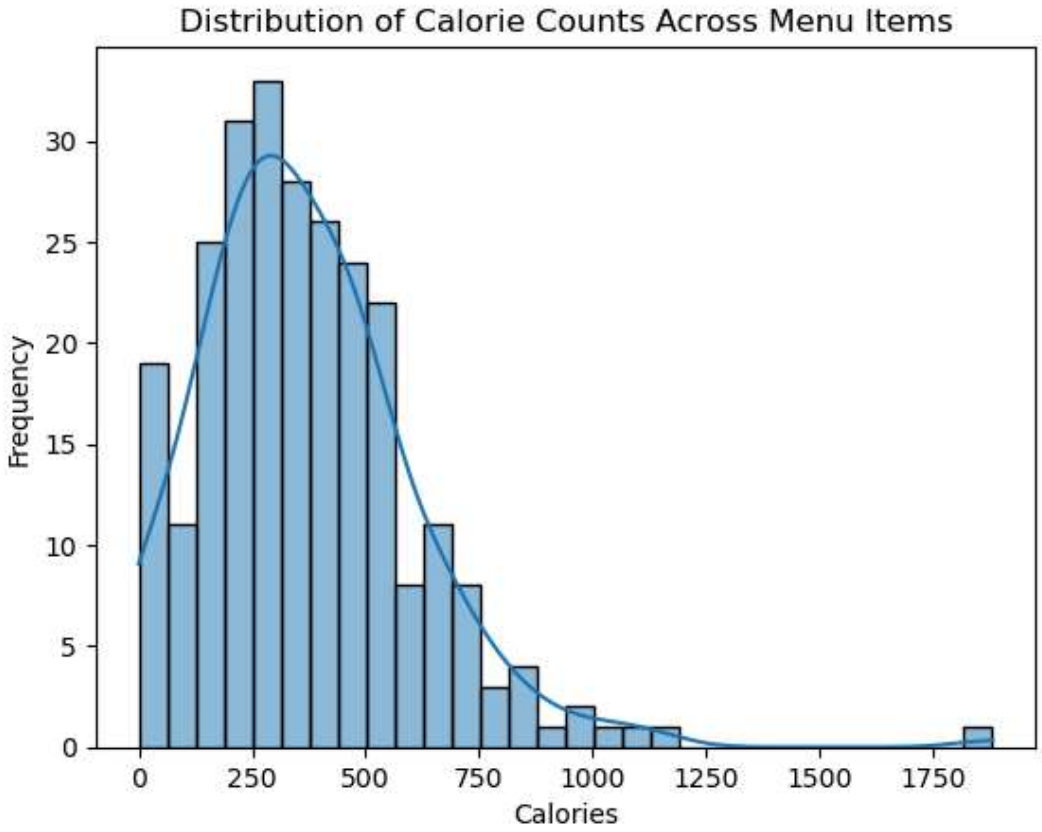
Analyze the distribution of calorie counts across menu items.

In [18]:
```
df['Calories'].value_counts()
```

Out[18]:
```
Calories
0      16
340    10
430    10
280     9
250     8
       ..
640     1
800     1
740     1
620     1
810     1
Name: count, Length: 78, dtype: int64
```

In [19]:
```
import seaborn as sns
```

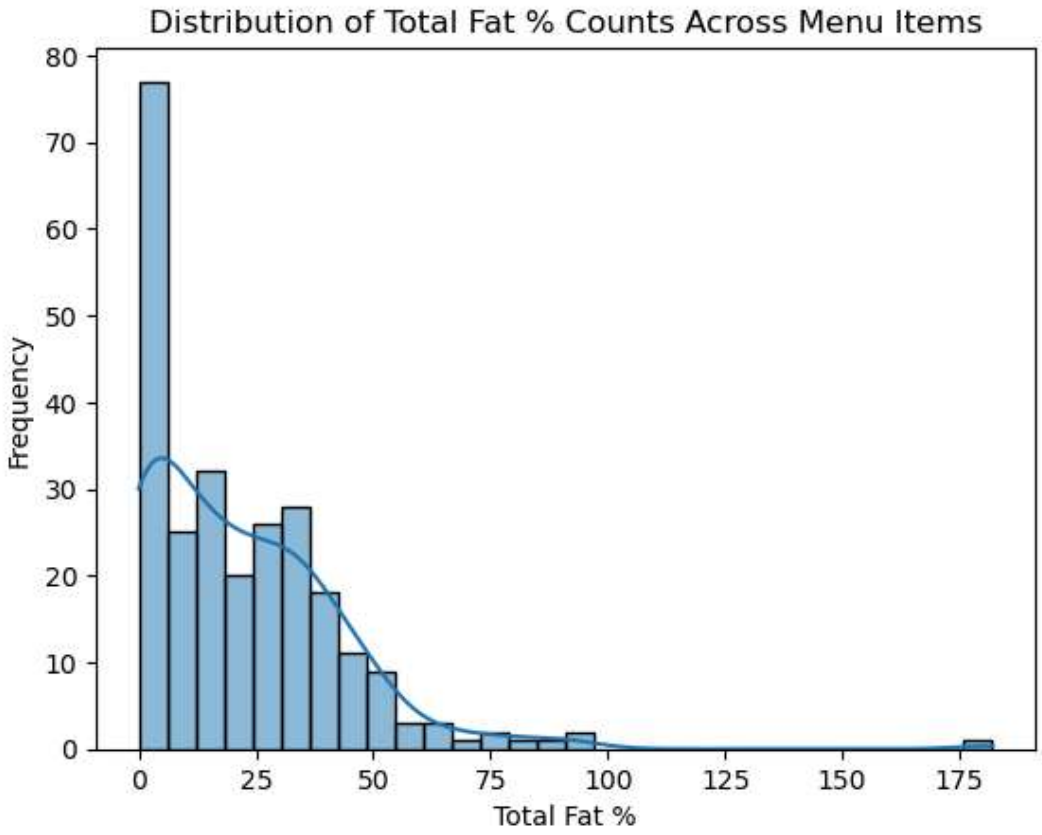
In [20]:
```
sns.histplot(df['Calories'], bins=30, kde=True)
plt.title('Distribution of Calorie Counts Across Menu Items')
plt.xlabel('Calories')
plt.ylabel('Frequency')
plt.show()
```



So here from the above histplot we see the calories between 200 to 500 is more distributed among the menu items and the plot is Right skewed indicates the calories mean is more sensitive toward the right as the outliers are towards the right.

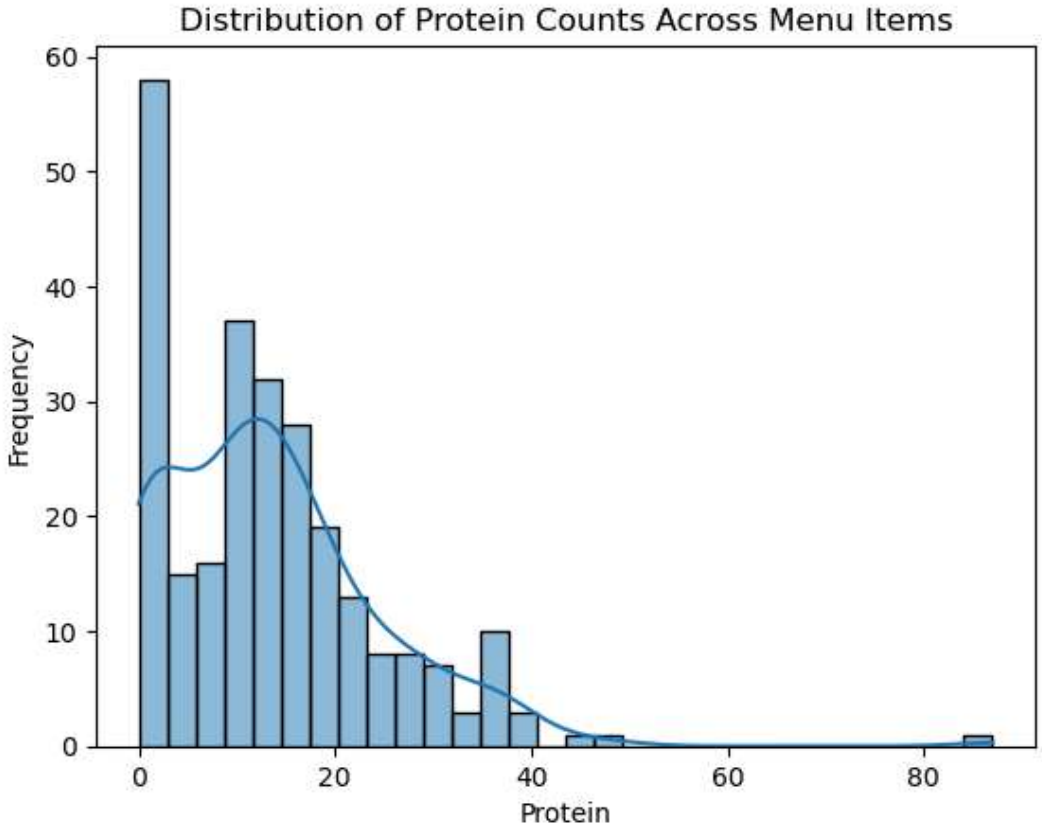
Explore the nutritional content (e.g., fat, protein, carbohydrates) of different items.

In [21]:
```
sns.histplot(df['Total Fat (% Daily Value)'], bins=30, kde=True)
plt.title('Distribution of Total Fat % Counts Across Menu Items')
plt.xlabel('Total Fat %')
plt.ylabel('Frequency')
plt.show()
```

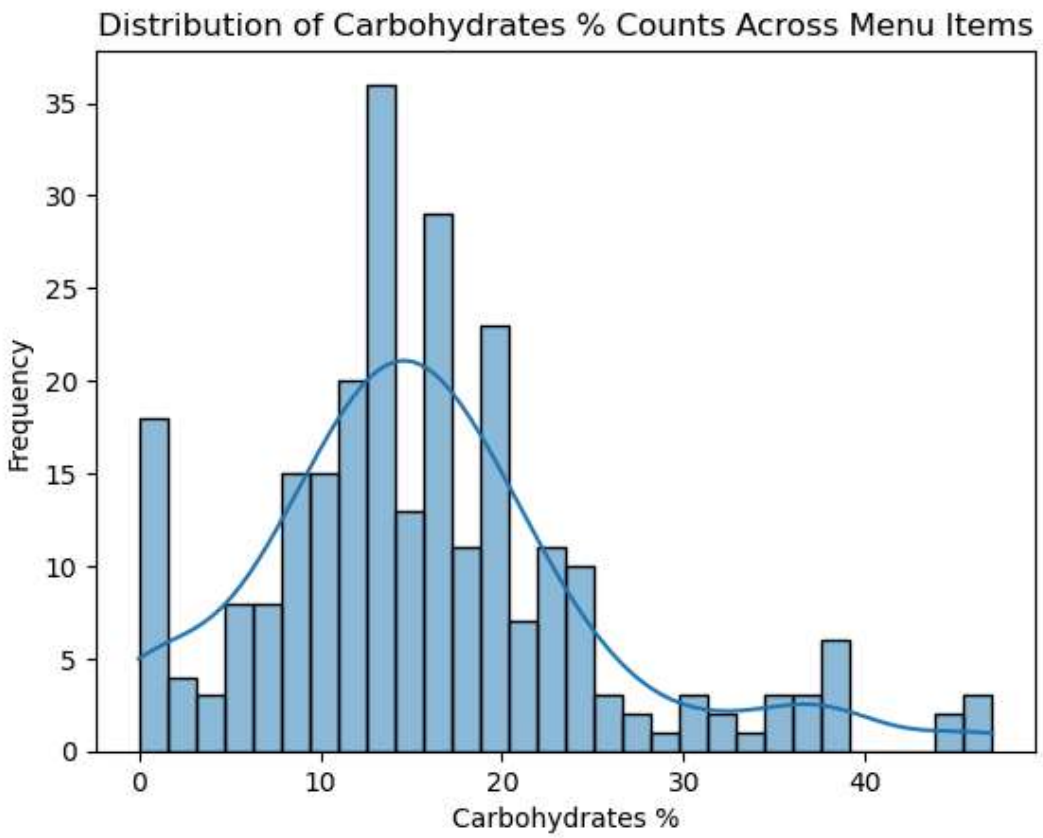


Total Fat content in menu item is baised and Right skewed indicates the fat mean is more sensitive toward the right as the outliers are towards the right. So we can conclude that in few items the fat content is high.

In [182]:
```
sns.histplot(df['Protein'], bins=30, kde=True)
plt.title('Distribution of Protein Counts Across Menu Items')
plt.xlabel('Protein')
plt.ylabel('Frequency')
plt.show()
```



In [23]:
```
sns.histplot(df['Carbohydrates (% Daily Value)'], bins=30, kde=True)
plt.title('Distribution of Carbohydrates % Counts Across Menu Items')
plt.xlabel('Carbohydrates %')
plt.ylabel('Frequency')
plt.show()
```



from the above bell curve we understand that the Carbohydrate distribution is quite uniform in all of the menu items. infact Carbohydrate value range between 10-20 is more in menu items.

In [113]:
```
nutritional_columns = ['Total Fat (% Daily Value)', 'Saturated Fat (% Daily Value)', '
                       'Sodium (% Daily Value)', 'Carbohydrates (% Daily Value)', 'Diet
```

In [134]:
```
nutritional_columns
```

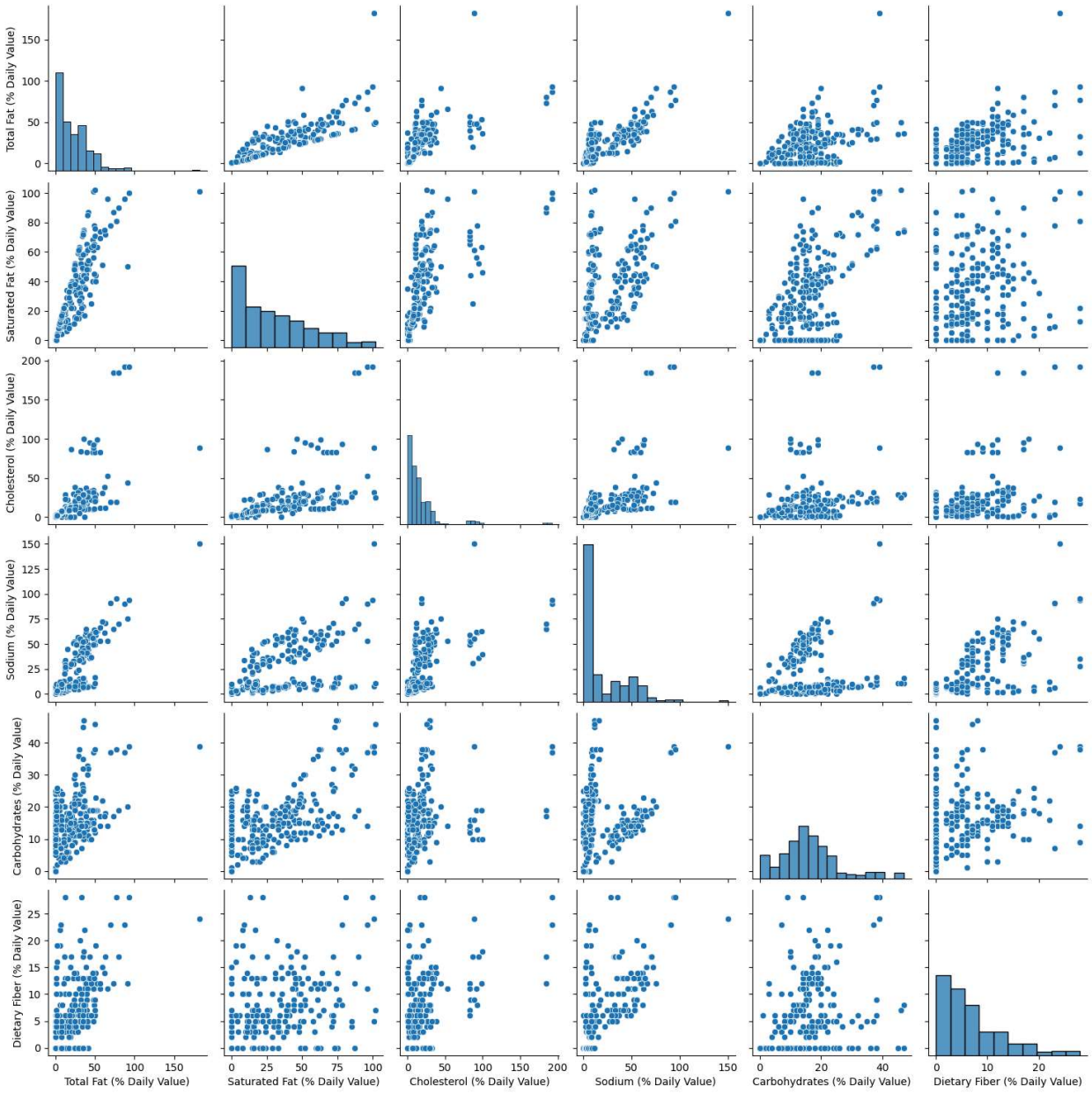
Out[134]:
```
['Total Fat (% Daily Value)',
 'Saturated Fat (% Daily Value)',
 'Cholesterol (% Daily Value)',
 'Sodium (% Daily Value)',
 'Carbohydrates (% Daily Value)',
 'Dietary Fiber (% Daily Value)']
```

Identify trends and patterns in the dataset.

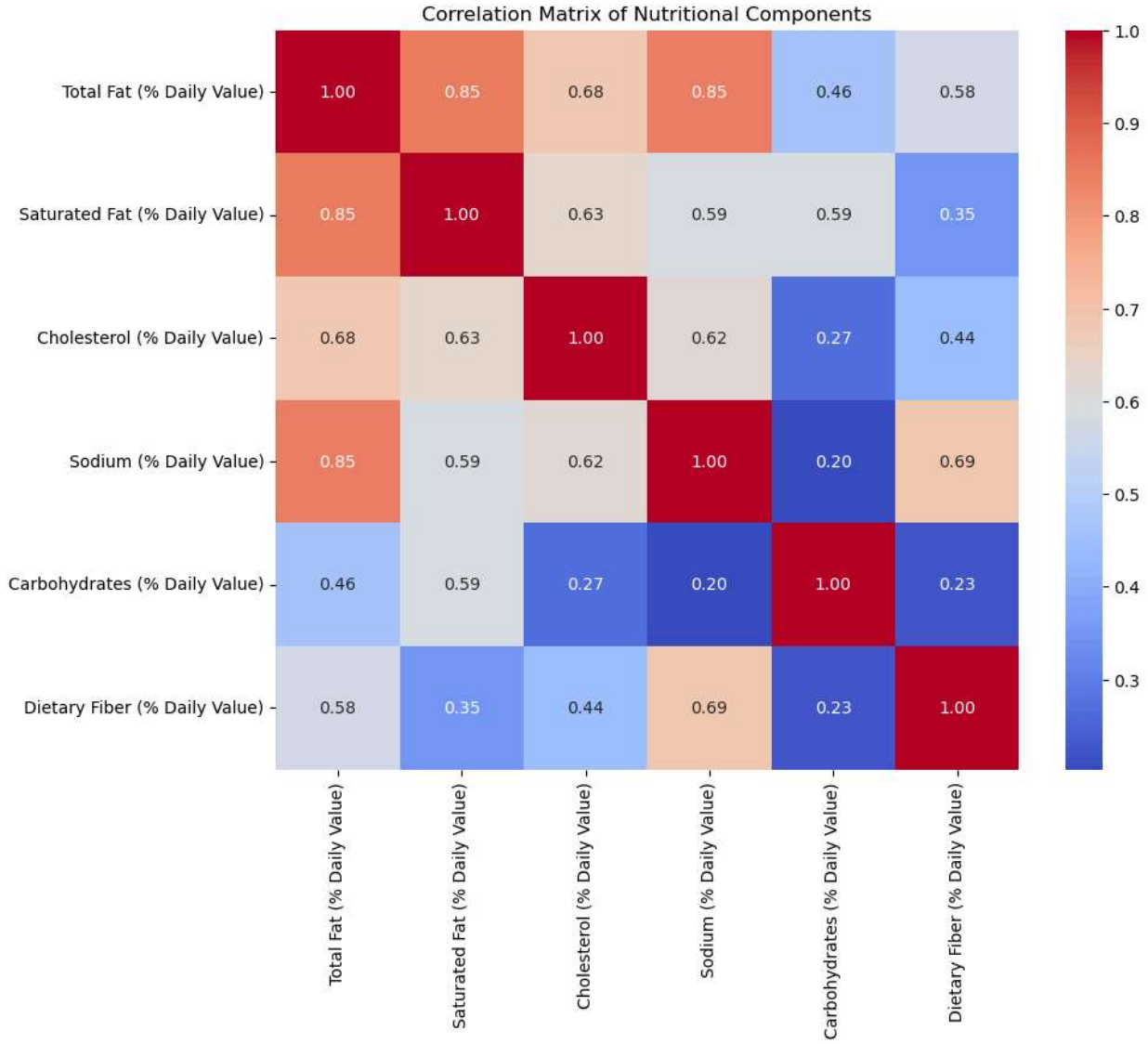
In [115]:
```
sns.pairplot(df[nutritional_columns])
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The
figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

Out[115]: <seaborn.axisgrid.PairGrid at 0x1a20d61a5d0>



In [116]:
```
correlation_matrix = df[nutritional_columns].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of Nutritional Components')
plt.show()
```



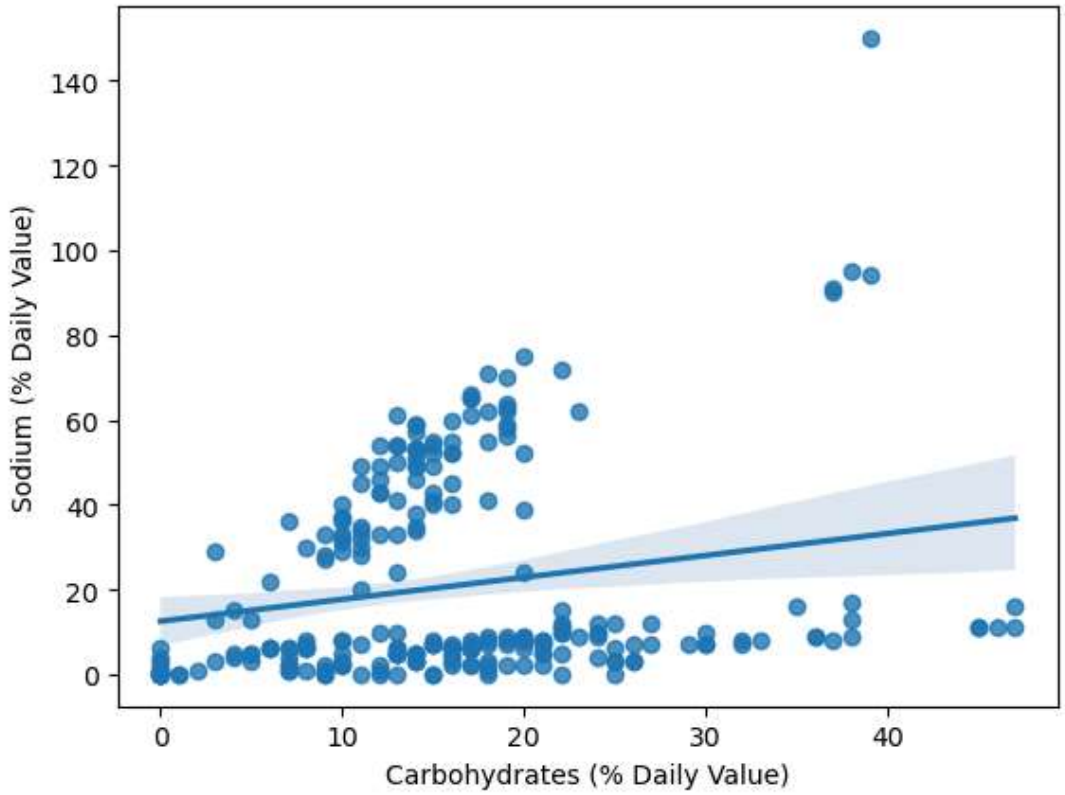
Here in the correlation matrix we check the relationship between two variables by checking the value of correlation coefficient. here the correlation coefficient between (Satured fat and Total fat), (sodium and total fat) are close to 1, indicating strong correlation between the nutients, which indicates that if satured fat increases in any menu item then the total the fat will also increases(as expected).

on the other hand the correlation between Carbohyderate and sodium is weak - 0.20(close to zero), indicating there is no relationship between these nutrient.

lets check the same below on regplot.

In [200]:
```
sns.regplot(data = df, x = 'Carbohydrates (% Daily Value)', y = 'Sodium (% Daily Value
```

Out[200]:
```
<Axes: xlabel='Carbohydrates (% Daily Value)', ylabel='Sodium (% Daily Value)'>
```

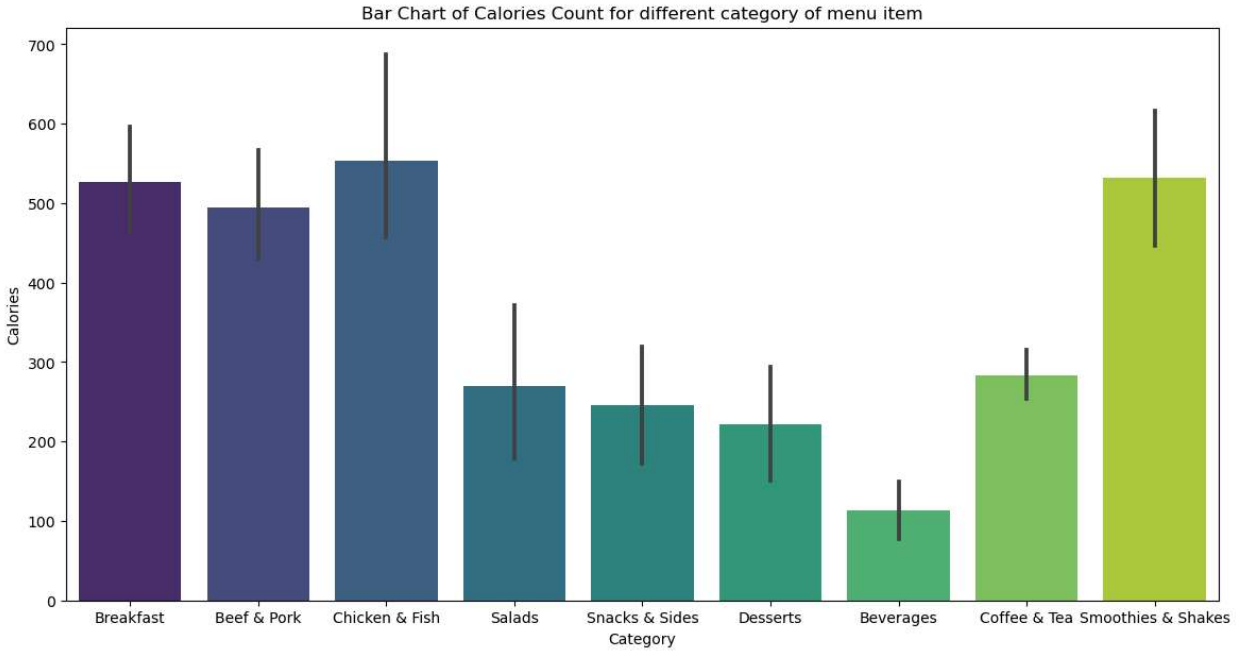


The above correlation coefficient value (0.20) for carbohydrates and sodium is weak and there is no linear regression relationship between these nutrients.

4. Data Visualization:

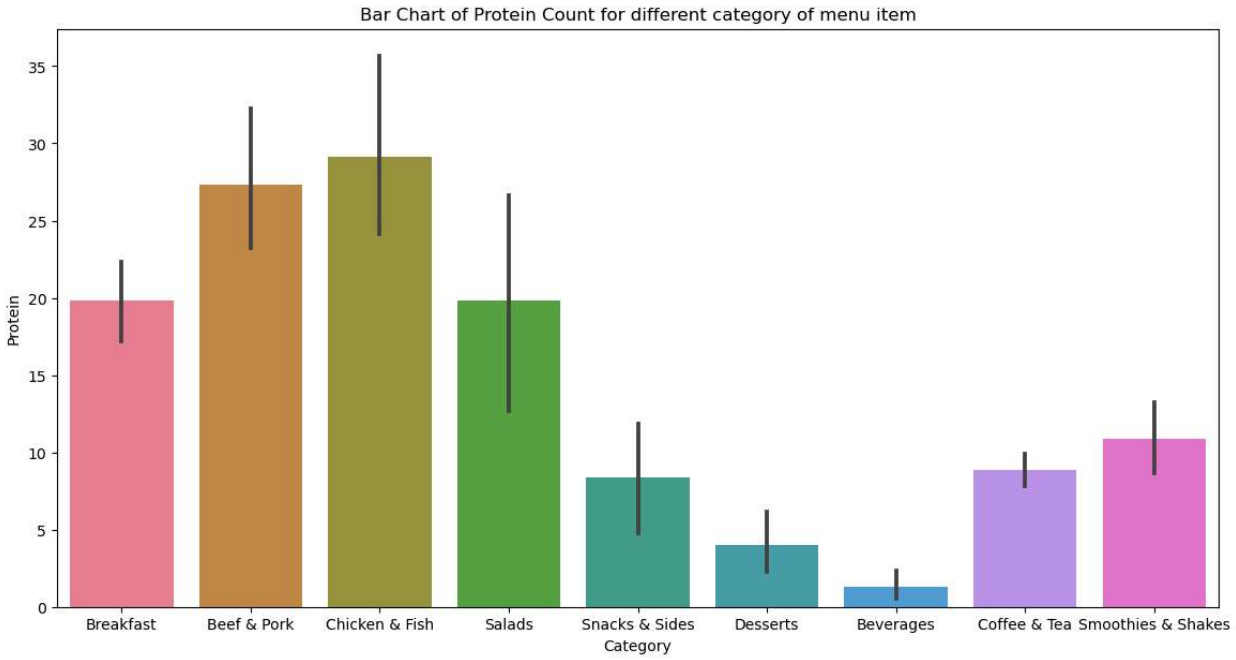
Create bar charts, histograms, and box plots to visualize calorie distribution and nutritional content. Compare nutritional characteristics of different food categories (e.g., burgers, salads, desserts).

In [117]:
```
plt.figure(figsize=(14, 7))
sns.barplot(x="Category", y="Calories", data=df, palette='viridis')
plt.title('Bar Chart of Calories Count for different category of menu item')
plt.xlabel('Category')
plt.ylabel('Calories')
plt.show()
```



Here we can see the chicken n fish have the highest calories and Beverages have lowest Calories

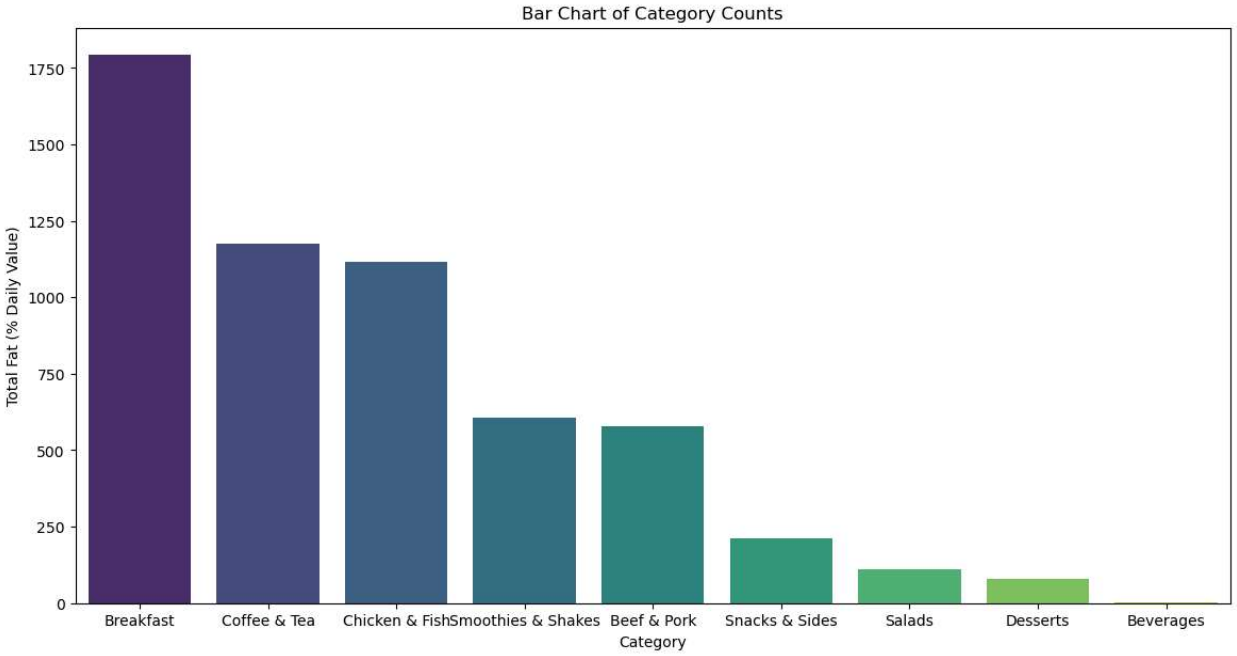
In [201]:
```
plt.figure(figsize=(14, 7))
sns.barplot(x="Category", y="Protein", data=df, palette='husl')
plt.title('Bar Chart of Protein Count for different category of menu item')
plt.xlabel('Category')
plt.ylabel('Protein')
plt.show()
```



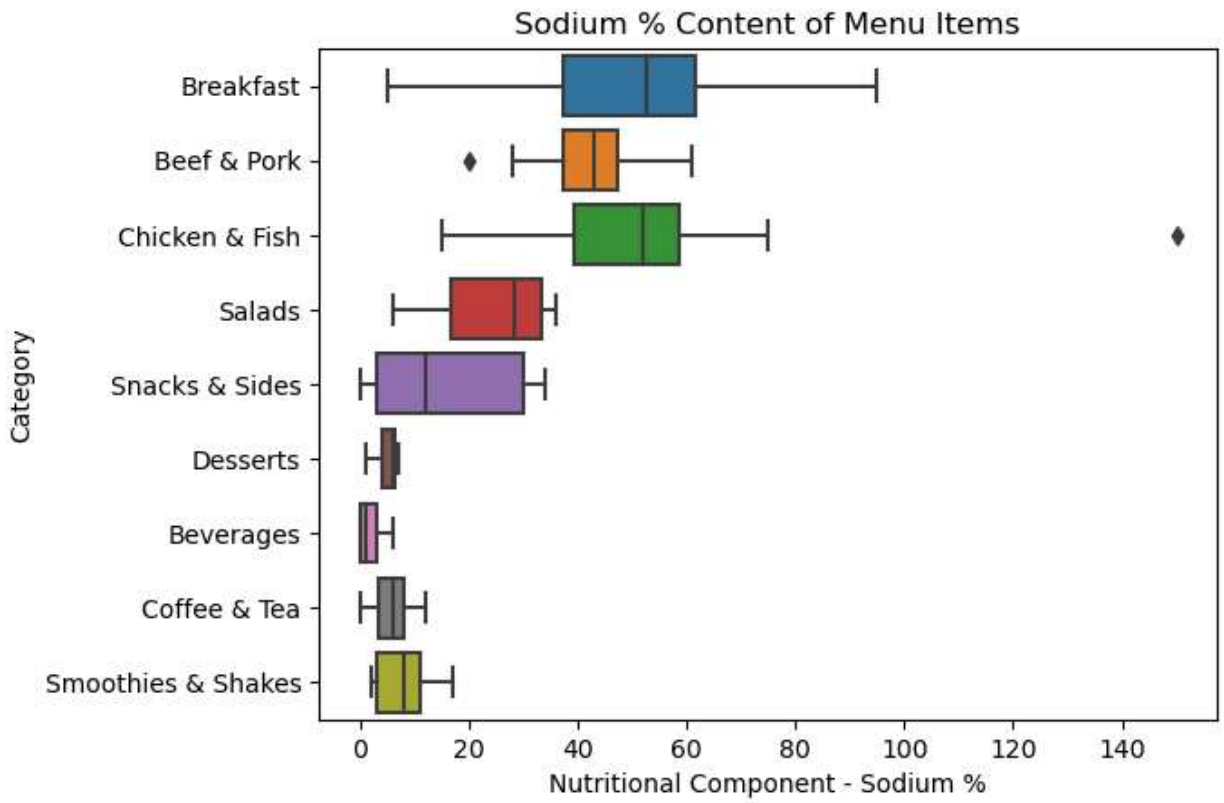
Well known fact - Chicken and fish have highest content of protein

In [93]:
```
fat = df.groupby(['Category'], as_index=False)['Total Fat (% Daily Value)'].sum().sort
```

In [98]:
```
plt.figure(figsize=(14, 7))
sns.barplot(x="Category", y="Total Fat (% Daily Value)", data=fat, palette='viridis')
plt.title('Bar Chart of Category Counts')
plt.xlabel('Category')
plt.ylabel('Total Fat (% Daily Value)')
plt.show()
```

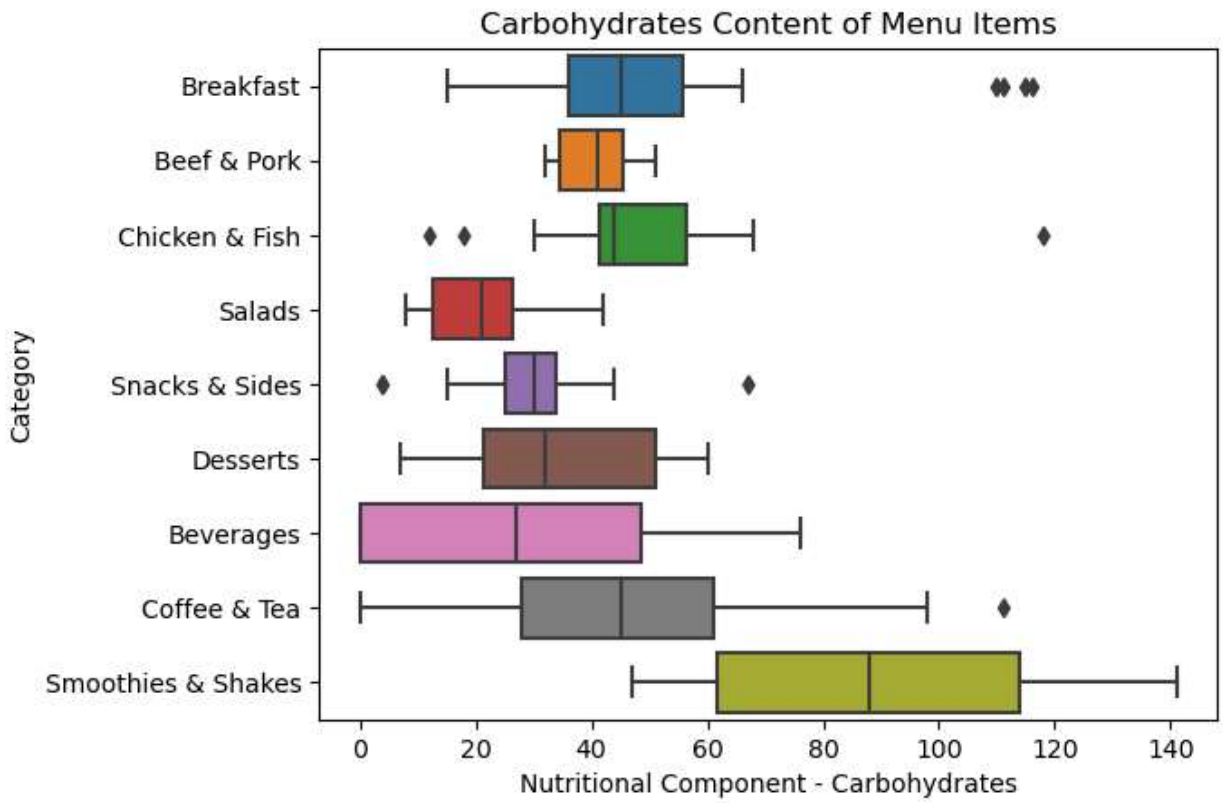


In [167]:
```
sns.boxplot(data=df, x='Sodium (% Daily Value)', y ='Category', orient='h')
plt.title('Sodium % Content of Menu Items')
plt.xlabel('Nutritional Component - Sodium %')
plt.ylabel('Category')
plt.show()
```



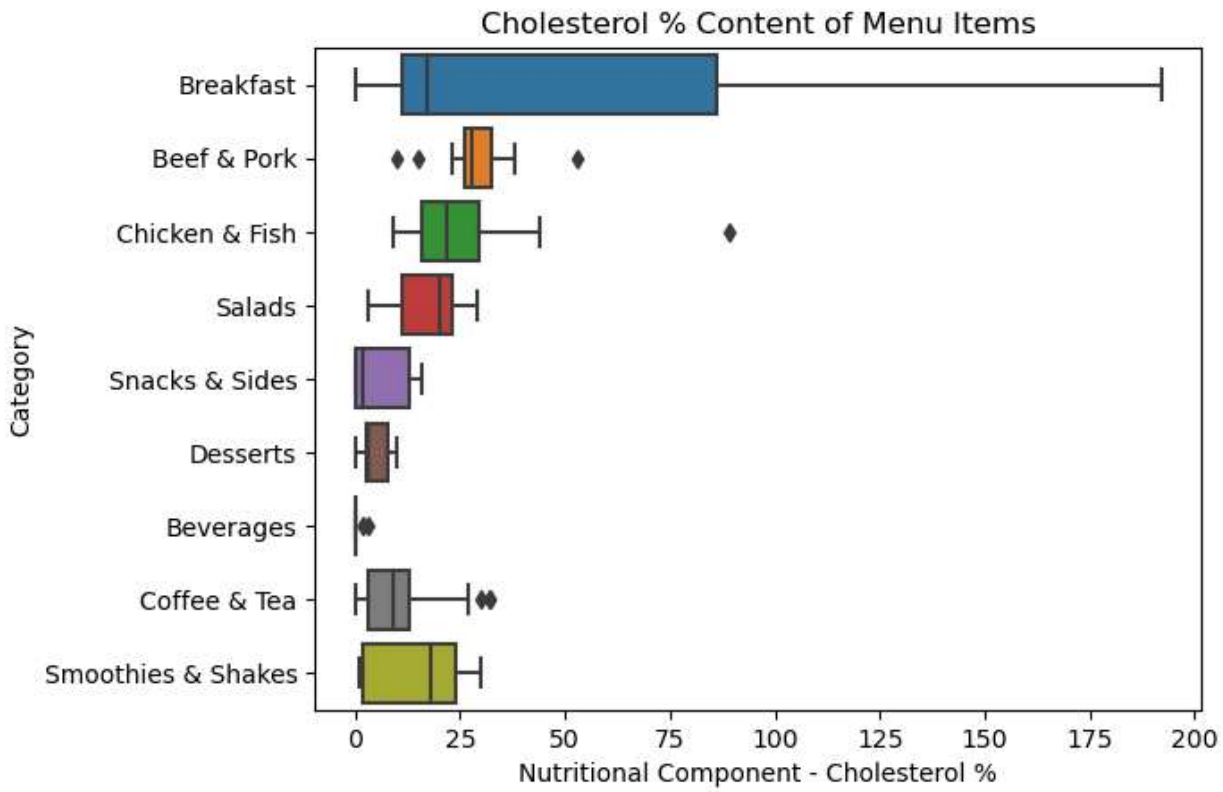
Well as expected, from the above boxplot sodium content is very less in Desserts and Beverges, but high sodium content in Snacks and Sides(expected this too).

In [121]:
```
sns.boxplot(data=df, x='Carbohydrates', y ='Category', orient='h')
plt.title('Carbohydrates Content of Menu Items')
plt.xlabel('Nutritional Component - Carbohydrates')
plt.ylabel('Category')
plt.show()
```



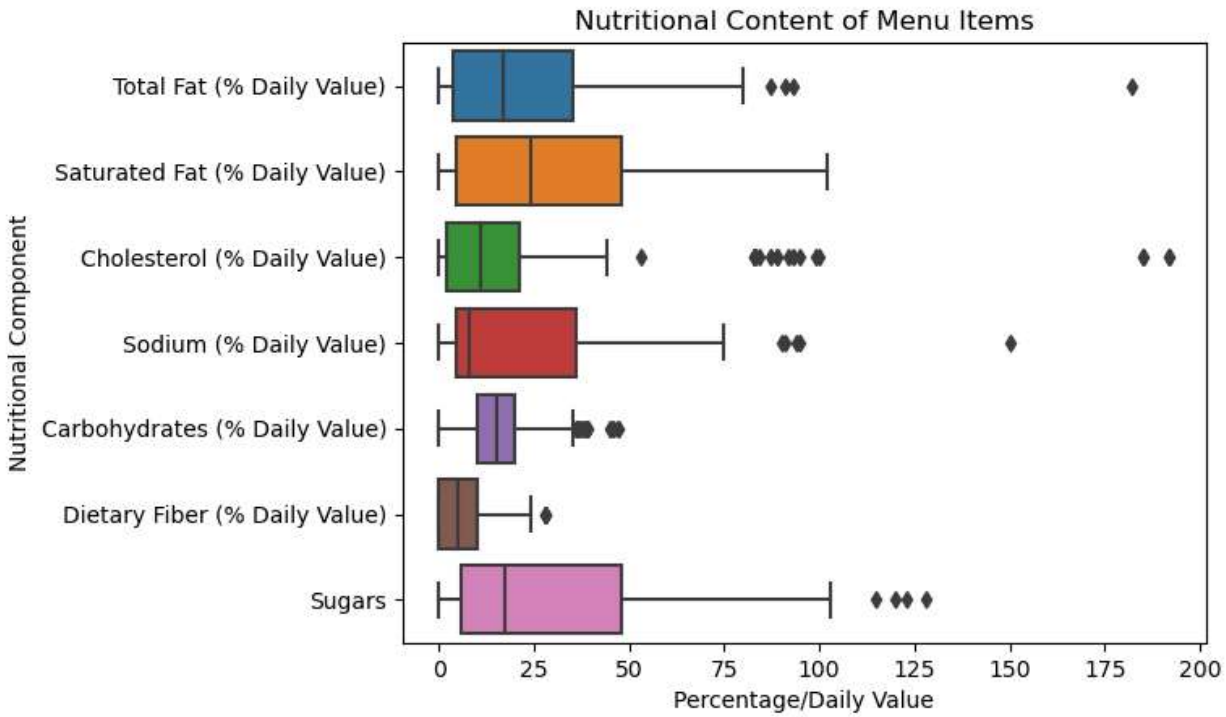
From the above boxplot we understand the content of carbohydrates is high in smoothies & shakes and Beverages, as compared with other nutrient ,every food category have whiskers and outliers indicating the carbohydrates is more dispersed among all food category except snacks and sides.

In [173]:
```
sns.boxplot(data=df, x='Cholesterol (% Daily Value)', y ='Category', orient='h')
plt.title('Cholesterol % Content of Menu Items')
plt.xlabel('Nutritional Component - Cholesterol %')
plt.ylabel('Category')
plt.show()
```

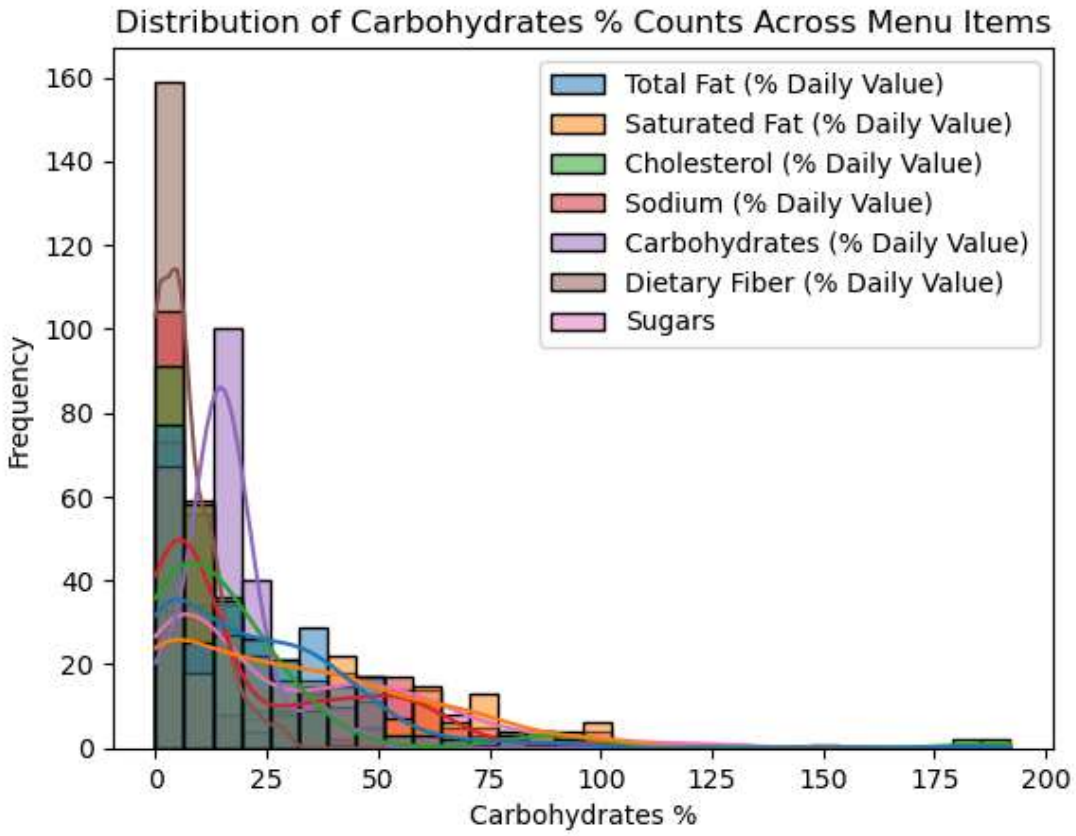


From above we can easily see that the content of Cholestrol is very high in breakfast, with long tail of maximum Whiskers value, which is not good for customers to have hight cholestrol intake in morning breakfast. Beverges have almost negligible content of Cholestrol which is a known fact.

In [74]:
```
sns.boxplot(data=df[nutritional_columns], orient='h')
plt.title('Nutritional Content of Menu Items')
plt.xlabel('Percentage/Daily Value')
plt.ylabel('Nutritional Component')
plt.show()
```



In [112]:
```
sns.histplot(df[nutritional_columns], bins=30, kde=True)
plt.title('Distribution of Carbohydrates % Counts Across Menu Items')
plt.xlabel('Carbohydrates %')
plt.ylabel('Frequency')
plt.show()
```



5. Nutrition-Based Insights:

Identify menu items with the highest and lowest calorie counts. Determine the average nutritional content of popular menu categories.

In [122]:
```
#Identify menu items with the highest and lowest calorie counts.
```

In [123]:
```
highest_cal = df.loc[df['Calories'].idxmax()]
```

In [131]:
```
highest_cal
```

Out[131]:
```
Category                                      Chicken & Fish
Item                             Chicken McNuggets (40 piece)
Serving Size                                 22.8 oz (646 g)
Calories                                                1880
Calories from Fat                                       1060
Total Fat                                              118.0
Total Fat (% Daily Value)                                182
Saturated Fat                                           20.0
Saturated Fat (% Daily Value)                            101
Trans Fat                                                1.0
Cholesterol                                              265
Cholesterol (% Daily Value)                               89
Sodium                                                  3600
Sodium (% Daily Value)                                   150
Carbohydrates                                            118
Carbohydrates (% Daily Value)                             39
Dietary Fiber                                              6
Dietary Fiber (% Daily Value)                             24
Sugars                                                     1
Protein                                                   87
Vitamin A (% Daily Value)                                  0
Vitamin C (% Daily Value)                                 15
Calcium (% Daily Value)                                    8
Iron (% Daily Value)                                      25
Name: 82, dtype: object
```

In [125]:
```
# ANS - Chicken McNuggets (40 piece) has highest calories = 1880
```

In [126]:
```
lowest_cal = df.loc[df['Calories'].idxmin()]
```

In [132]:
```
lowest_cal
```

Out[132]:
```
Category                                 Beverages
Item                              Diet Coke (Small)
Serving Size                           16 fl oz cup
Calories                                          0
Calories from Fat                                 0
Total Fat                                       0.0
Total Fat (% Daily Value)                         0
Saturated Fat                                   0.0
Saturated Fat (% Daily Value)                     0
Trans Fat                                       0.0
Cholesterol                                       0
Cholesterol (% Daily Value)                       0
Sodium                                           10
Sodium (% Daily Value)                            0
Carbohydrates                                     0
Carbohydrates (% Daily Value)                     0
Dietary Fiber                                     0
Dietary Fiber (% Daily Value)                     0
Sugars                                            0
Protein                                           0
Vitamin A (% Daily Value)                         0
Vitamin C (% Daily Value)                         0
Calcium (% Daily Value)                           0
Iron (% Daily Value)                              0
Name: 114, dtype: object
```

In [128]:
```
# ANS - Diet Coke (Small) has lowest calories - 0
```

Determine the average nutritional content of popular menu categories.

In [193]:
```
VitaminA = df.groupby(['Category'], as_index=False)['Vitamin A (% Daily Value)'].mean(
```

In [194]:
```
VitaminA
```

Out[194]:

	Category	Vitamin A (% Daily Value)
6	Salads	146.666667
3	Chicken & Fish	20.444444
7	Smoothies & Shakes	18.750000
4	Coffee & Tea	10.736842
0	Beef & Pork	6.933333
2	Breakfast	6.928571
5	Desserts	5.142857
8	Snacks & Sides	4.846154
1	Beverages	0.740741

In [195]:
```
Iron = df.groupby(['Category'], as_index=False)['Iron (% Daily Value)'].mean().sort_va
```

In [196]:
```
Iron
```

Out[196]:

	Category	Iron (% Daily Value)
0	Beef & Pork	23.333333
2	Breakfast	17.142857
3	Chicken & Fish	16.370370
6	Salads	10.333333
8	Snacks & Sides	6.615385
5	Desserts	4.000000
7	Smoothies & Shakes	3.964286
4	Coffee & Tea	2.147368
1	Beverages	0.296296

In [197]:
```
dex=False)['Calcium (% Daily Value)'].mean().sort_values(by='Calcium (% Daily Value)',
```

In [198]:
```
Calcium
```

Out[198]:

	Category	Calcium (% Daily Value)
7	Smoothies & Shakes	35.571429
4	Coffee & Tea	28.294737
0	Beef & Pork	23.000000
2	Breakfast	16.166667
3	Chicken & Fish	15.444444
6	Salads	15.000000
5	Desserts	11.428571
8	Snacks & Sides	6.000000
1	Beverages	2.962963

6. Documentation and Reporting:

Summarize the findings and insights from the analysis.

1. From the barplot we identify that Chicken and fish have highest content of protein so they are high calorie food items which is good and healthy to have.
2. Carbohydrate content is good in smoothies & shakes & in Beverges, its good for refereshment with good amount of nutrient value.

3. Cholestrol or fat content is quite high in breakfast, which is not good for customers to have high cholestrol intake in morning breakfast, unhealthy.
4. Sodium content is high in Snacks and sides which is absolutey fine and it is minimally distributed among the other food items as per the requirement, not too high which is good. Sodium intake should always be minimal.
5. Salads have good content of VitaminA, very healthy.
6. Calcium and Iron content is quite low.
7. From the correlation matrix we understand that the correlation between Carbohyderate and sodium is weak(close to zero), indicating there is no relationship between these nutrient. and correlation between saturated fat and total fat is strong - 0.83 (close to 1) indicates that the indicates that if

In [ ]:

Explain how the nutritional analysis could benefit McDonald's customers and the organization.

McDonald's is everyones favourite place to go for children, young or elderly people, so Nutritional analysis is important that provides customers the information about the food they are consuming, nutritional content of menu items such as calorie counts, fat, protein, cholestrol, carbohydrates, sodium, and many more.

Today in the world of adulteration, people are very conscious about their intake. Growing health problems making them health-conscious, by having high level understanding of the nutritional composition of their meals, customers can make informed decisions about their food choices, helping them select options that align with their dietary preferences, health goals, and nutritional needs.

It gives complete transparency and confidence to the customers about their choices that enhances their trust and customer satisfaction.

It improve the marketing strategy by giving clear insight about its food, promoting social awareness about the nutritional content in their food items, that strengthens its reputation in the market.

Product development - The most crucial part of this analysis , if they found some food items have less nutrition content, or something unhealthy not good for health, then they could analyze and make right decision to improvise the quality of the food items.

In [ ]: