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

In [2]:
```
df = pd.read_excel('fmcg.xlsx')
```

In [3]:
```
df.head()
```

Out[3]:

| | Company | Revenues ($MM) | Profit ($MM) | Profit as % of Revenues | Category |
|---|---|---|---|---|---|
| **0** | Johnson & Johnson | 71890 | 18540 | 0.257894 | Pharmaceuticals |
| **1** | Procter & Gamble | 71726 | 10508 | 0.146502 | Household & Personal Products |
| **2** | Pepsico | 62789 | 6329 | 0.100798 | Food |
| **3** | Pfizer | 52824 | 7215 | 0.136586 | Pharmaceuticals |
| **4** | Coca-Cola | 41863 | 6527 | 0.155913 | Beverages |

In [4]:
```
# Define revenue thresholds for categorical representation
small_threshold = 20000
medium_threshold = 50000

# Categorize companies based on revenue thresholds
def categorize_company_size(revenue):
    if revenue < small_threshold:
        return 'Small'
    elif revenue < medium_threshold:
        return 'Medium'
    else:
        return 'Large'

# Create new feature 'Company_Size_Categorical'
df['Company_Size_Categorical'] = df['Revenues ($MM)'].apply(categorize_compa
```

In [5]:
```
df.head()
```

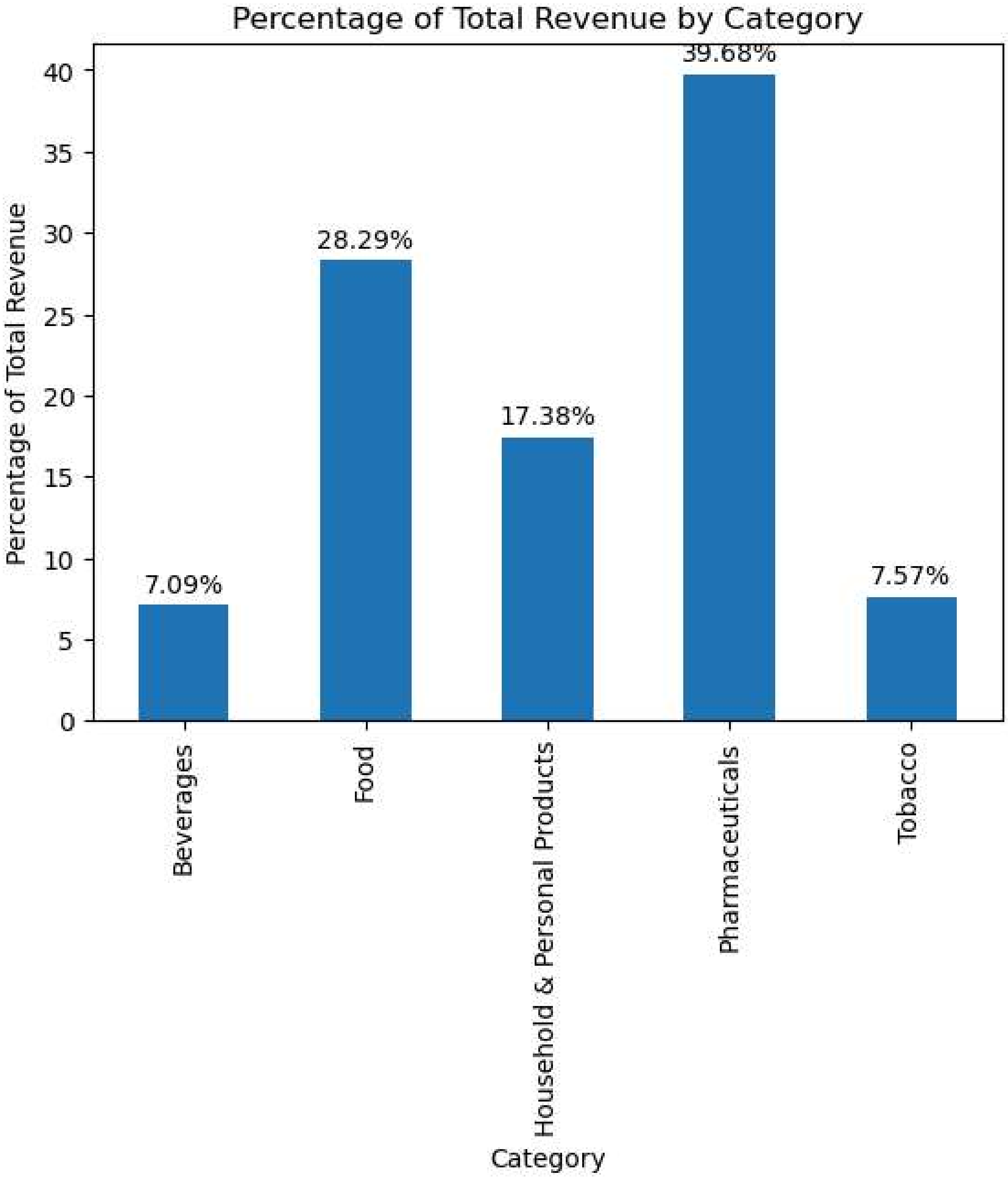
Out[5]:

| | Company | Revenues ($MM) | Profit ($MM) | Profit as % of Revenues | Category | Company_Size_Categorical |
|---|---|---|---|---|---|---|
| **0** | Johnson & Johnson | 71890 | 18540 | 0.257894 | Pharmaceuticals | Large |
| **1** | Procter & Gamble | 71726 | 10508 | 0.146502 | Household & Personal Products | Large |
| **2** | Pepsico | 62789 | 6329 | 0.100798 | Food | Large |
| **3** | Pfizer | 52824 | 7215 | 0.136586 | Pharmaceuticals | Large |
| **4** | Coca-Cola | 41863 | 6527 | 0.155913 | Beverages | Medium |

## Find the sum of revenue belonging to specific category

In [6]:

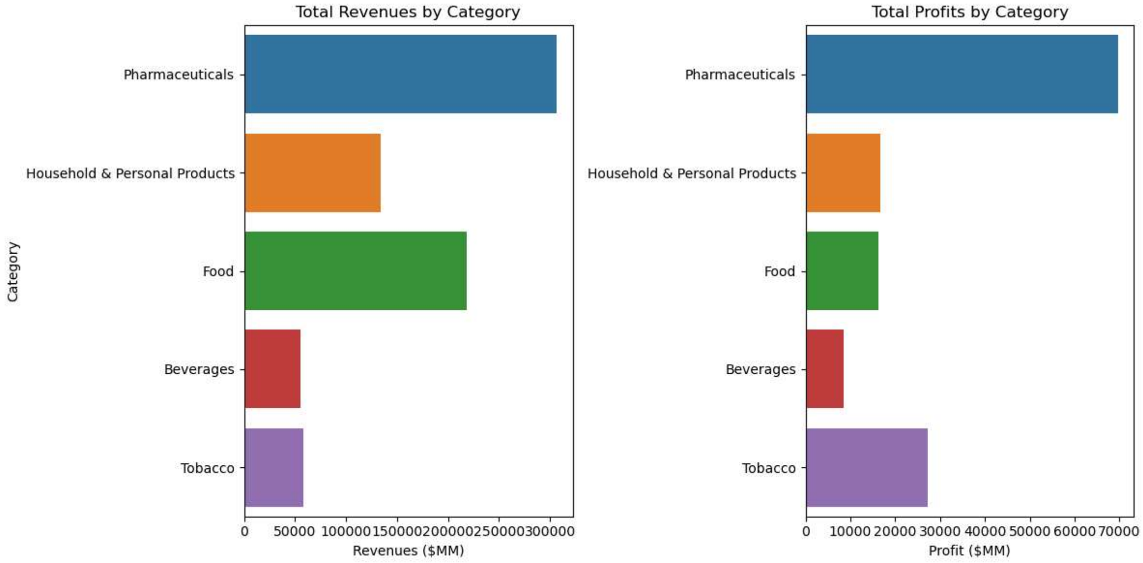
```
# Calculate total revenue
total_revenue = df['Revenues ($MM)'].sum()

# Group by category and calculate sum of revenues
category_revenue = df.groupby('Category')['Revenues ($MM)'].sum()

# Calculate percentage of revenue for each category
percentage_revenue = (category_revenue / total_revenue) * 100

# Plot the bar plot
ax = percentage_revenue.plot(kind='bar')

# Add data labels as percentages
for i in ax.patches:
    ax.text(i.get_x() + i.get_width() / 2, i.get_height() + 0.5, f'{i.get_he

# Add labels and title
plt.xlabel('Category')
plt.ylabel('Percentage of Total Revenue')
plt.title('Percentage of Total Revenue by Category')

# Show plot
plt.show()
```


Percentage of Total Revenue by Category
Percentage of Total Revenue
0
5
10
15
20
25
30
35
40
7.09%
28.29%
17.38%
39.68%
7.57%
Beverages
Food
Household & Personal Products
Pharmaceuticals
Tobacco
Category

## How do revenues and profits vary across different industries?

In [7]:
```
# Set up the plot
plt.figure(figsize=(12, 6))

# Create side-by-side bar plots for revenues and profits by category
plt.subplot(1, 2, 1)
sns.barplot(data=df, x='Revenues ($MM)', y='Category', estimator=sum, ci=Nor
plt.title('Total Revenues by Category')
plt.xlabel('Revenues ($MM)')
plt.ylabel('Category')

plt.subplot(1, 2, 2)
sns.barplot(data=df, x='Profit ($MM)', y='Category', estimator=sum, ci=None)
plt.title('Total Profits by Category')
plt.xlabel('Profit ($MM)')
plt.ylabel('')

plt.tight_layout()
plt.show()
```

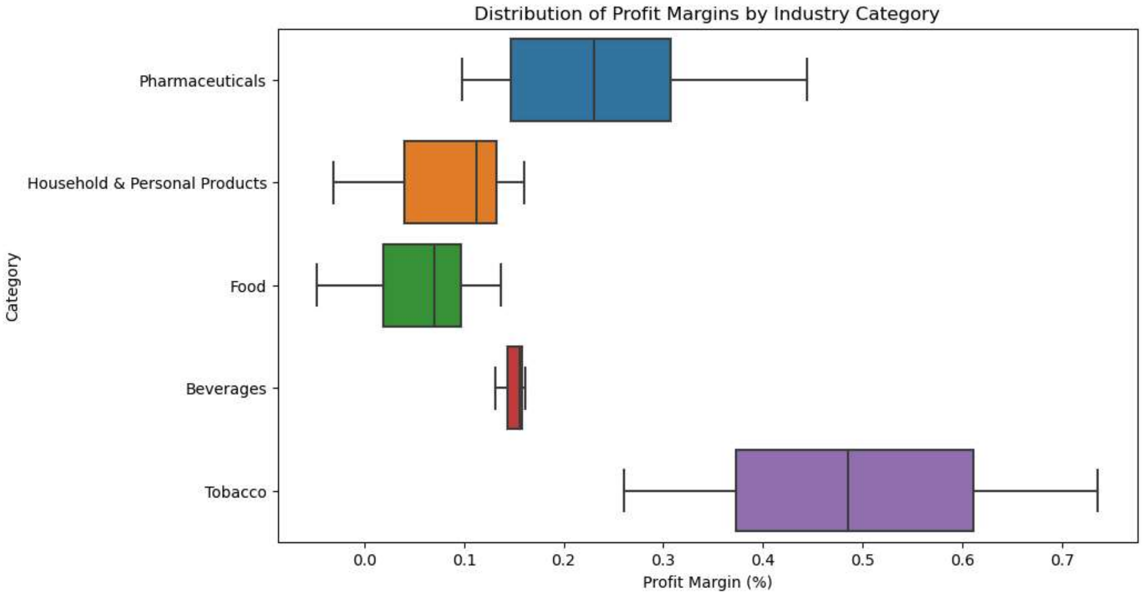
```
C:\Users\mohan\AppData\Local\Temp\ipykernel_16524\1699647112.py:6: FutureW
arning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(data=df, x='Revenues ($MM)', y='Category', estimator=sum, ci
=None)
C:\Users\mohan\AppData\Local\Temp\ipykernel_16524\1699647112.py:12: Future
Warning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(data=df, x='Profit ($MM)', y='Category', estimator=sum, ci=N
one)
```



## What is the distribution of profit margins within each industry?

In [8]:

```
# Set up the plot
plt.figure(figsize=(10, 6))

# Create box plot for profit margins by category
sns.boxplot(data=df, x='Profit as % of Revenues', y='Category')
plt.title('Distribution of Profit Margins by Industry Category')
plt.xlabel('Profit Margin (%)')
plt.ylabel('Category')

plt.show()
```



### How do the top companies in terms of revenue compare to the top companies in terms of profit?

In [9]:

```
# Sort the DataFrame by revenue and profit in descending order
df_sorted_revenue = df.sort_values(by='Revenues ($MM)', ascending=False)
df_sorted_profit = df.sort_values(by='Profit ($MM)', ascending=False)

# Get the top companies in terms of revenue and profit
top_companies_revenue = df_sorted_revenue.head(5)
top_companies_profit = df_sorted_profit.head(5)

# Merge the two sets of top companies to identify overlap
top_companies_combined = pd.merge(top_companies_revenue, top_companies_profi

# Display the combined DataFrame
top_companies_combined
```

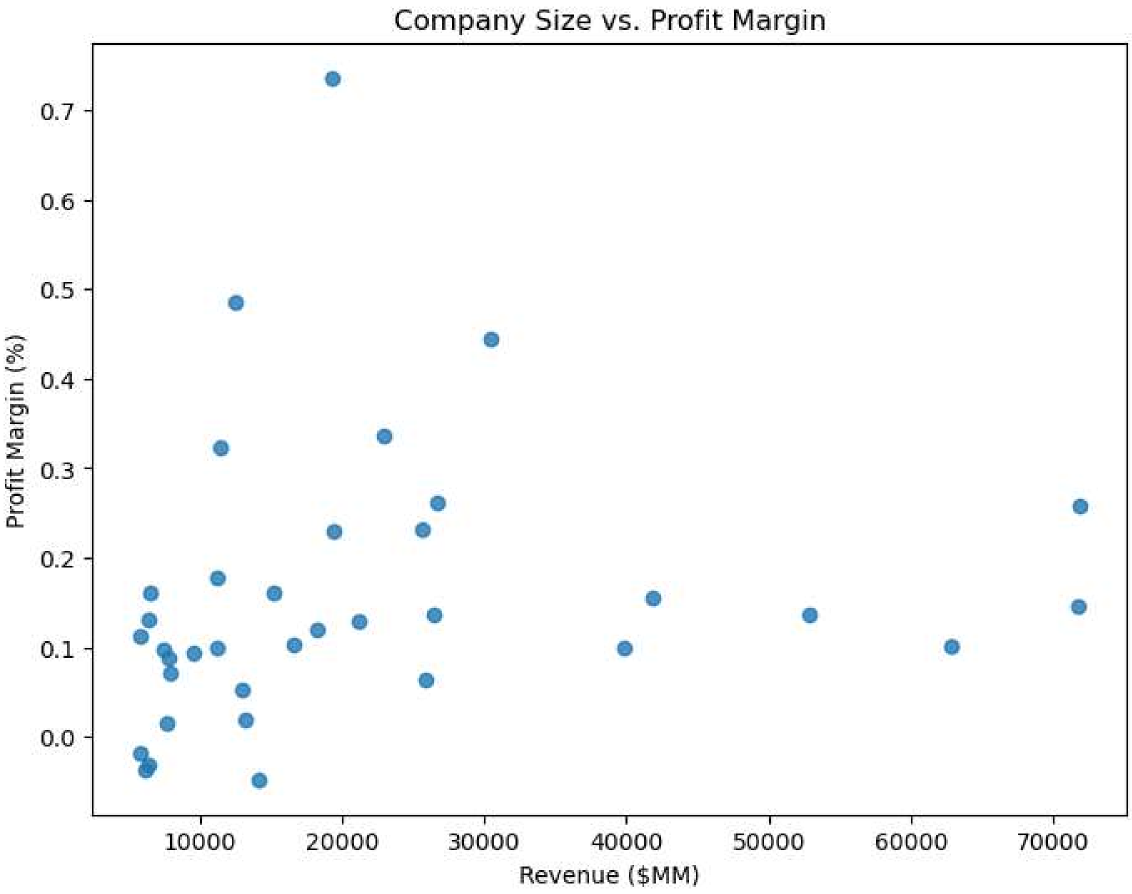
Out[9]:

| | Company | Revenues ($MM)_revenue | Profit ($MM)_revenue | Profit as % of Revenues_revenue | Category_revenue | Company |
|---|---|---|---|---|---|---|
| **0** | Johnson & Johnson | 71890 | 18540 | 0.257894 | Pharmaceuticals | |
| **1** | Procter & Gamble | 71726 | 10508 | 0.146502 | Household & Personal Products | |

### Is there a relationship between company size (measured by revenue) and profit margin?

In [10]:
```
# Set up the plot
plt.figure(figsize=(8, 6))

# Create scatter plot
plt.scatter(df['Revenues ($MM)'], df['Profit as % of Revenues'], alpha=0.8)
plt.title('Company Size vs. Profit Margin')
plt.xlabel('Revenue ($MM)')
plt.ylabel('Profit Margin (%)')

# Calculate correlation coefficient
correlation_coefficient = df['Revenues ($MM)'].corr(df['Profit as % of Rever
print(f"Correlation Coefficient: {correlation_coefficient:.2f}")

plt.show()
```

```
Correlation Coefficient: 0.17
```



### What is the market share of each company within its respective industry?

In [11]:

```
# Calculate total revenue for each industry category
total_revenue_by_category = df.groupby('Category')['Revenues ($MM)'].sum()

# Calculate market share for each company within its industry category
df['Market Share'] = df.apply(lambda row: row['Revenues ($MM)'] / total_reve

# Display the DataFrame with market share
df[['Company', 'Category', 'Revenues ($MM)', 'Market Share']]
```

Out[11]:

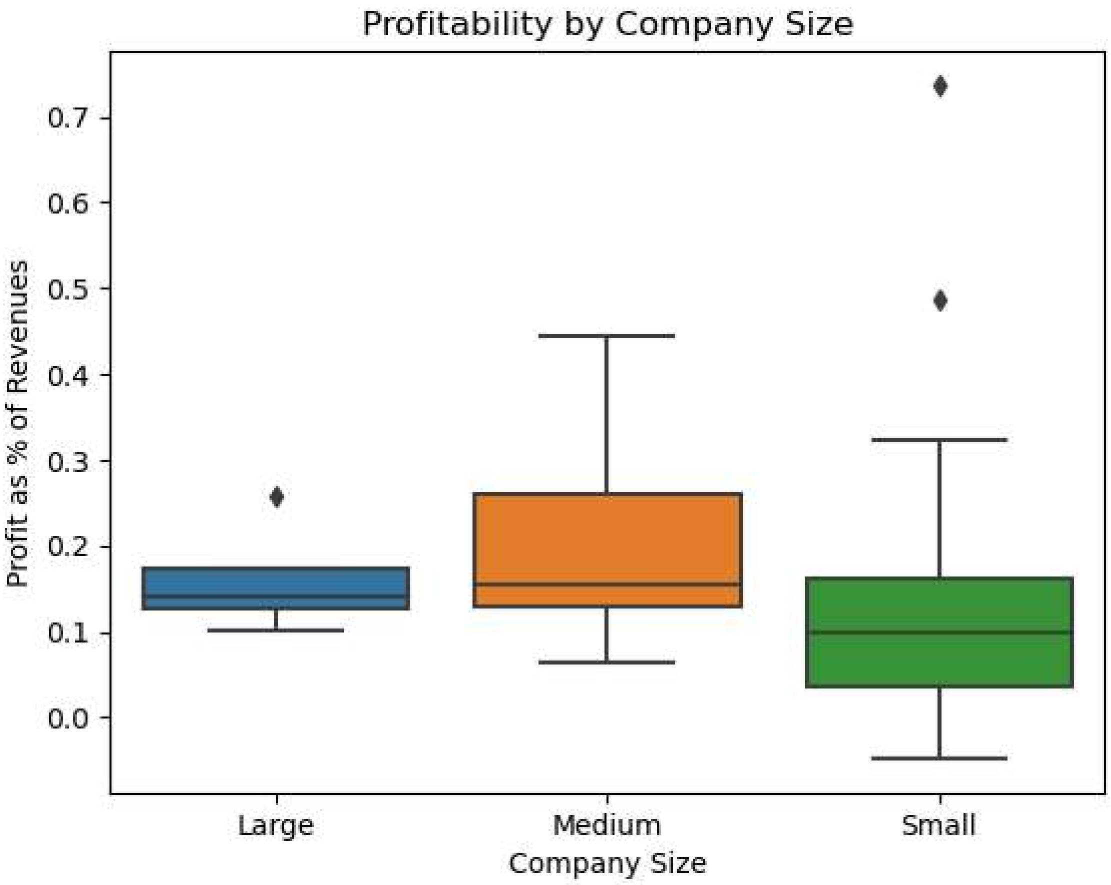
| | Company | Category | Revenues ($MM) | Market Share |
|---|---|---|---|---|
| **0** | Johnson & Johnson | Pharmaceuticals | 71890 | 0.234271 |
| **1** | Procter & Gamble | Household & Personal Products | 71726 | 0.533668 |
| **2** | Pepsico | Food | 62789 | 0.287018 |
| **3** | Pfizer | Pharmaceuticals | 52824 | 0.172140 |
| **4** | Coca-Cola | Beverages | 41863 | 0.763213 |
| **5** | Merck | Pharmaceuticals | 39807 | 0.129721 |
| **6** | Gilead Sciences | Pharmaceuticals | 30390 | 0.099033 |
| **7** | Philip Morris International | Tobacco | 26685 | 0.455959 |
| **8** | Kraft Heinz | Food | 26487 | 0.121076 |
| **9** | Mondelez International | Food | 25923 | 0.118498 |
| **10** | Abbvie | Pharmaceuticals | 25638 | 0.083548 |
| **11** | Amgen | Pharmaceuticals | 22991 | 0.074922 |
| **12** | Eli Lilly | Pharmaceuticals | 21222 | 0.069157 |
| **13** | Bristol-Myers Squibb | Pharmaceuticals | 19427 | 0.063308 |
| **14** | Altria Group | Tobacco | 19337 | 0.330406 |
| **15** | Kimberly-Clark | Household & Personal Products | 18202 | 0.135430 |
| **16** | General Mills | Food | 16563 | 0.075712 |
| **17** | Colgate-Palmolive | Household & Personal Products | 15195 | 0.113056 |
| **18** | Conagra Brands | Food | 14134 | 0.064609 |
| **19** | Land O'Lakes | Food | 13233 | 0.060490 |
| **20** | Kellogg | Food | 13014 | 0.059489 |
| **21** | Reynolds American | Tobacco | 12503 | 0.213635 |
| **22** | Biogen | Pharmaceuticals | 11449 | 0.037309 |
| **23** | Estee Lauder | Household & Personal Products | 11262 | 0.083793 |
| **24** | Celgene | Pharmaceuticals | 11229 | 0.036592 |
| **25** | Hormel Foods | Food | 9523 | 0.043531 |
| **26** | Campbell Soup | Food | 7961 | 0.036391 |
| **27** | J. M. Smucker | Food | 7811 | 0.035705 |
| **28** | Dean Foods | Food | 7710 | 0.035244 |
| **29** | Hershey | Food | 7440 | 0.034009 |
| **30** | Constellation Brands | Beverages | 6548 | 0.119378 |
| **31** | Dr. Pepper Snapple Group | Beverages | 6440 | 0.117409 |
| **32** | HRG Group | Household & Personal Products | 6403 | 0.047641 |
| **33** | Treehouse Foods | Food | 6175 | 0.028227 |
| **34** | Avon Products | Household & Personal Products | 5853 | 0.043548 |
| **35** | Clorox | Household & Personal Products | 5761 | 0.042864 |

In [12]:

```
df.to_csv('fmcg_main.csv')
```

In [13]:

```
df=pd.read_csv('fmcg_main.csv')
# Plotting the relationship between company size and profitability
sns.boxplot(x='Company_Size_Categorical', y='Profit as % of Revenues', data=
plt.title('Profitability by Company Size')
plt.xlabel('Company Size')
plt.ylabel('Profit as % of Revenues')
plt.show()
```



In [14]:

```
import pandas as pd
from scipy.stats import f_oneway



# Extract profitability data for each company size category
small = df[df['Company_Size_Categorical'] == 'Small']['Profit as % of Revenu
medium = df[df['Company_Size_Categorical'] == 'Medium']['Profit as % of Reve
large = df[df['Company_Size_Categorical'] == 'Large']['Profit as % of Revenu

# Perform ANOVA test
statistic, p_value = f_oneway(small, medium, large)

# Set significance level
alpha = 0.05

# Print results
print("ANOVA Test Results:")
print("Test Statistic:", statistic)
print("p-value:", p_value)

# Interpret results
if p_value < alpha:
    print("There are significant differences in profitability among companie
else:
    print("There are no significant differences in profitability among compa
```

```
ANOVA Test Results:
Test Statistic: 0.62917886057582
p-value: 0.5393020854290076
There are no significant differences in profitability among companies of d
ifferent sizes.
```

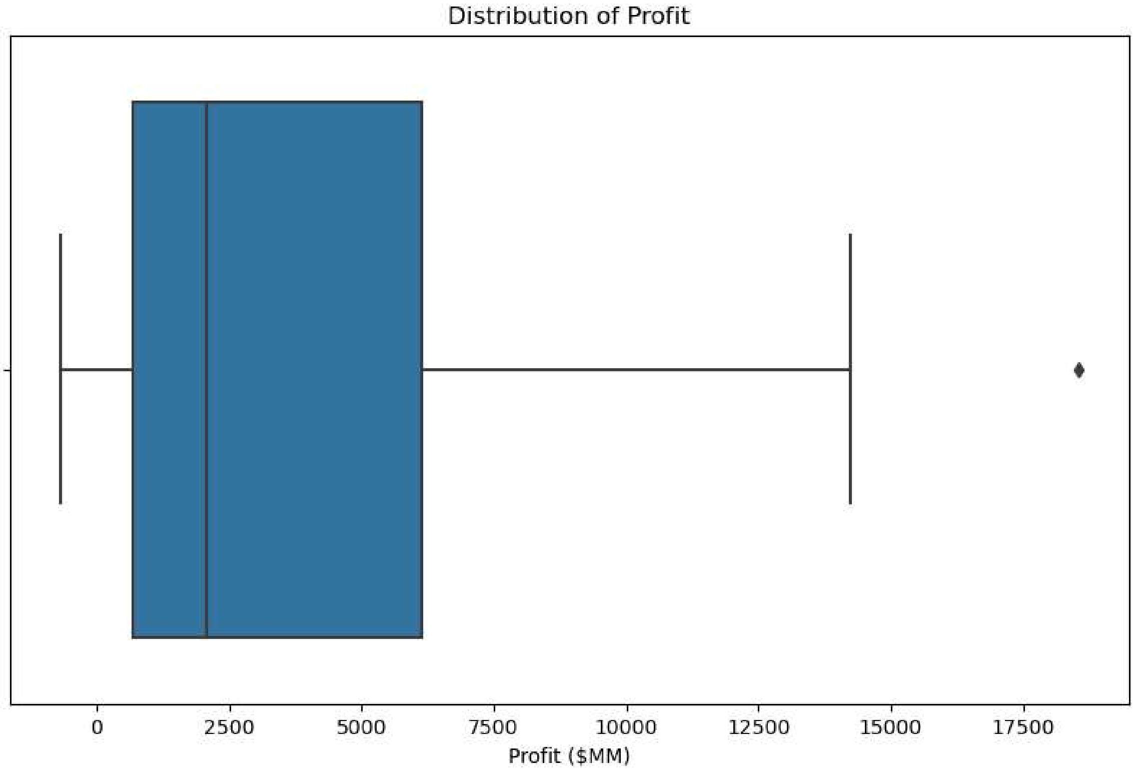
In [15]:

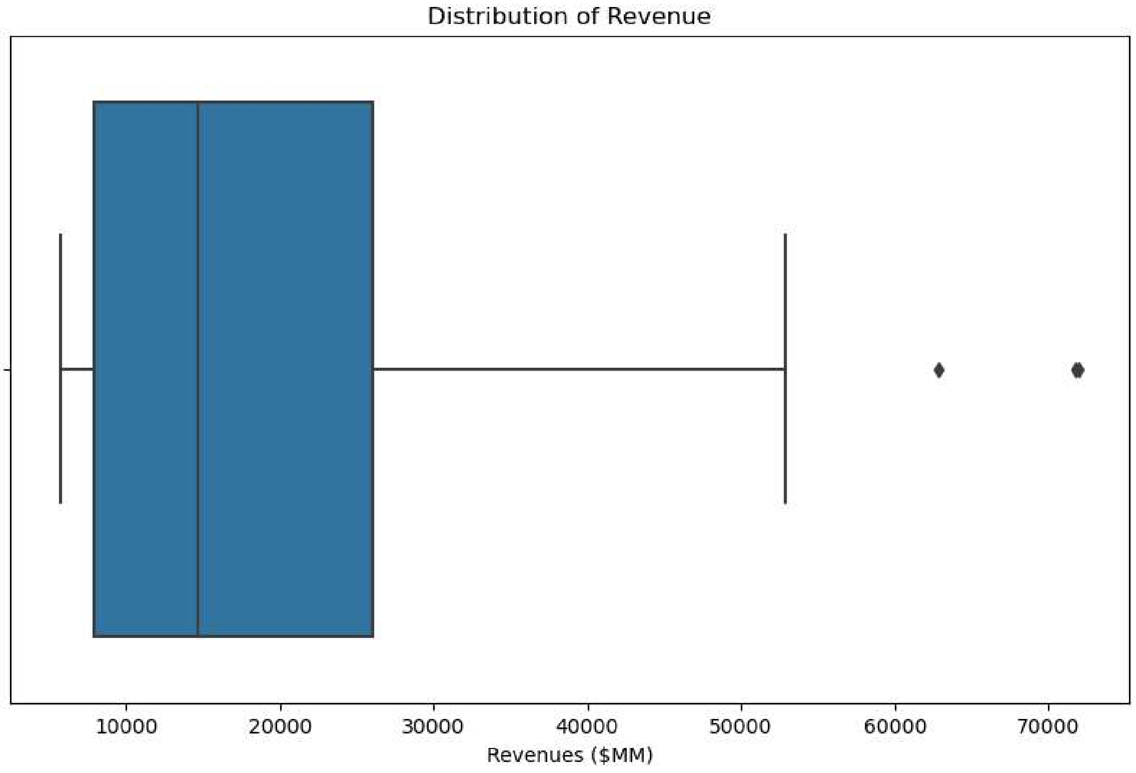
```


# Calculate average profitability for each category
average_profitability = df.groupby('Category')['Profit as % of Revenues'].me

# Identify the most profitable category
most_profitable_category = average_profitability.idxmax()
average_profitability_value = average_profitability.max()

print("Most Profitable Category:", most_profitable_category)
print("Average Profitability:", average_profitability_value)
```

```
Most Profitable Category: Tobacco
Average Profitability: 0.4943889257543861
```

In [16]:

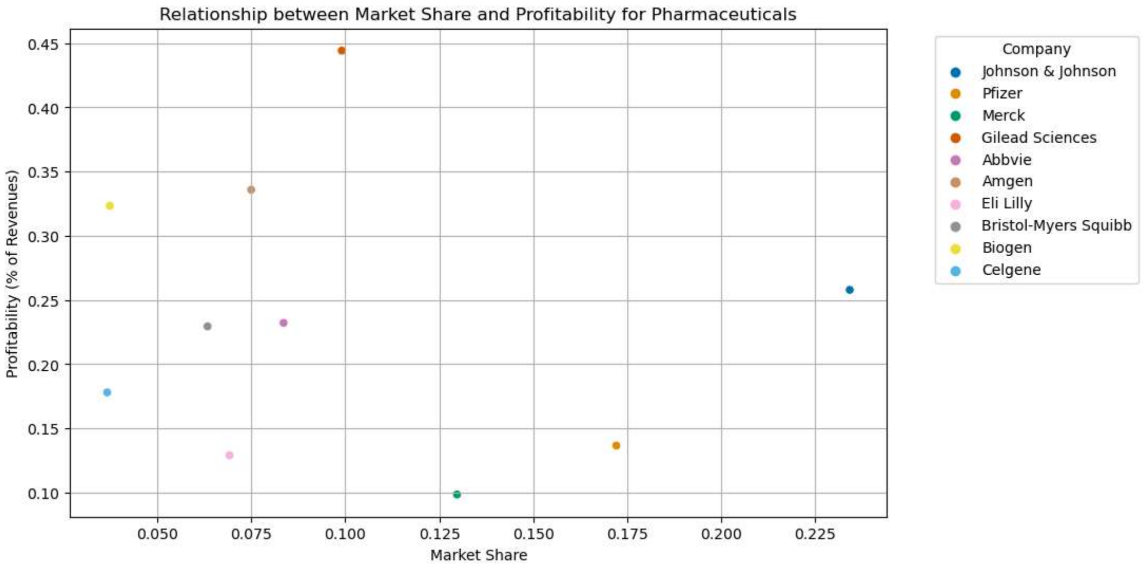
```
# Visualize distribution of profit using a box plot
plt.figure(figsize=(10, 6))
sns.boxplot(x='Profit ($MM)', data=df)
plt.title('Distribution of Profit')
plt.show()

# Visualize distribution of revenue using a box plot
plt.figure(figsize=(10, 6))
sns.boxplot(x='Revenues ($MM)', data=df)
plt.title('Distribution of Revenue')
plt.show()

# Calculate z-scores for profit and revenue
df['Profit Z-Score'] = (df['Profit ($MM)'] - df['Profit ($MM)'].mean()) / df
df['Revenue Z-Score'] = (df['Revenues ($MM)'] - df['Revenues ($MM)'].mean())

# Identify outliers based on z-scores (threshold of |z-score| > 3)
profit_outliers = df[abs(df['Profit Z-Score']) > 3]
revenue_outliers = df[abs(df['Revenue Z-Score']) > 3]

print("Profit Outliers:")
print(profit_outliers)
print("\nRevenue Outliers:")
print(revenue_outliers)
```





```
Profit Outliers:
   Unnamed: 0           Company  Revenues ($MM)  Profit ($MM)  \
0           0  Johnson & Johnson           71890         18540   

   Profit as % of Revenues         Category Company_Size_Categorical  \
0                 0.257894  Pharmaceuticals                    Large   

   Market Share  Profit Z-Score  Revenue Z-Score  
0      0.234271        3.242016         2.763337  

Revenue Outliers:
Empty DataFrame
Columns: [Unnamed: 0, Company, Revenues ($MM), Profit ($MM), Profit as % o
f Revenues, Category, Company_Size_Categorical, Market Share, Profit Z-Sco
re, Revenue Z-Score]
Index: []
```

In [17]:
```
# Select a specific category for analysis (e.g., Pharmaceuticals)
category = 'Pharmaceuticals'

# Filter data for the selected category
category_data = df[df['Category'] == category]

# Visualize the relationship between market share and profitability using a
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Market Share', y='Profit as % of Revenues', data=category
plt.title('Relationship between Market Share and Profitability for {}'.forma
plt.xlabel('Market Share')
plt.ylabel('Profitability (% of Revenues)')
plt.legend(title='Company', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()
```



In [ ]: