ANALYSIS FOR PYTHON PROJECT

TOPIC – LIST OF LARGEST COMPANIES IN THE UNITED STATES BY REVENUE

Submitted by: Rithika Anand

PGDM-BDA

045047

PROJECT OBJECTIVES

The project aims to design and implement a web scraping solution for data extraction from a selected website – https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue, followed by a comprehensive data analysis.

On the data analysis front, the project focuses on transforming the scraped data, conducting exploratory and statistical analysis. This analysis is supported by snippets.

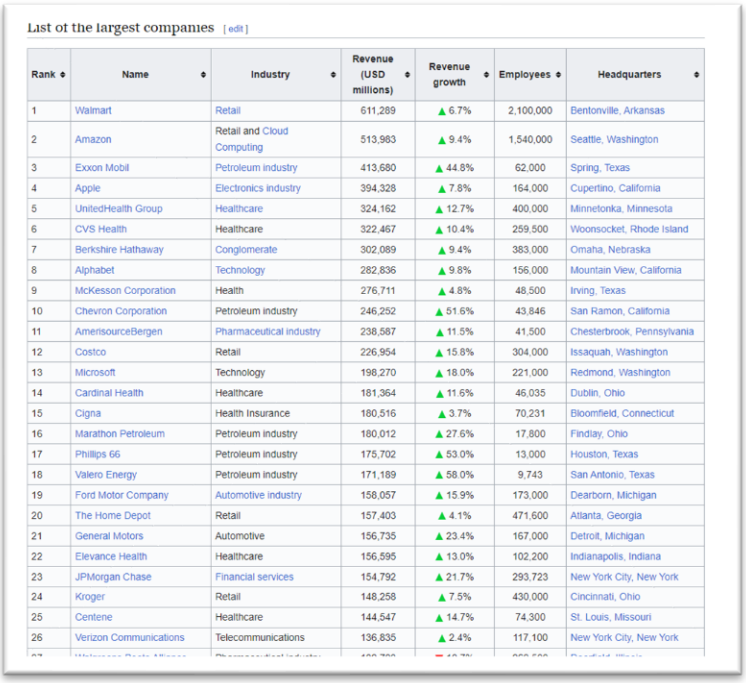
Ultimately, the project aims to provide valuable insights through a well-documented report based on the analysed data.

GENERAL DESCRIPTION OF DATA

The website chosen is - https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue.

The data available on this Wikipedia page **"List of largest companies in the United States by revenue"** typically includes a list of the largest companies in the United States, ranked as per their annual revenue. This data is organized into a table format on the webpage.

1. **Rankings:** The companies are usually ranked in order of their annual revenue, with the highest revenue company listed at the top.
2. **Revenue:** The most prominent data point is the annual revenue of each company, usually measured in billions of U.S. dollars.
3. **Industry Sector:** The industry or sector to which each company belongs is mentioned, providing information about the business area in which the company operates.
4. **Change in Rank:** In some tables, you might find information on how a company's ranking has changed compared to the previous year. This can provide insights into a company's performance over time.

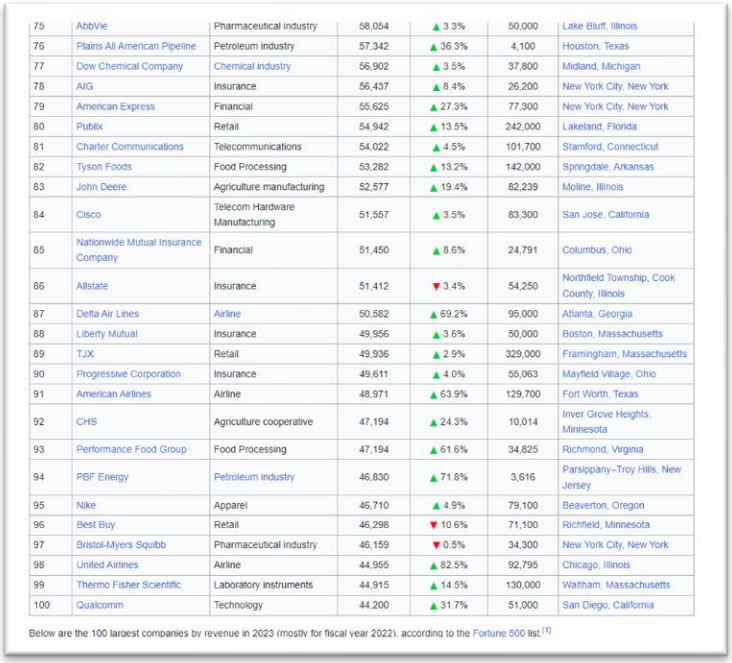


List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington
13	Microsoft	Technology	198,270	▲ 18.0%	221,000	Redmond, Washington
14	Cardinal Health	Healthcare	181,364	▲ 11.6%	46,035	Dublin, Ohio
15	Cigna	Health Insurance	180,516	▲ 3.7%	70,231	Bloomfield, Connecticut
16	Marathon Petroleum	Petroleum industry	180,012	▲ 27.6%	17,800	Findlay, Ohio
17	Phillips 66	Petroleum industry	175,702	▲ 53.0%	13,000	Houston, Texas
18	Valero Energy	Petroleum industry	171,189	▲ 58.0%	9,743	San Antonio, Texas
19	Ford Motor Company	Automotive industry	158,057	▲ 15.9%	173,000	Dearborn, Michigan
20	The Home Depot	Retail	157,403	▲ 4.1%	471,600	Atlanta, Georgia
21	General Motors	Automotive	156,735	▲ 23.4%	167,000	Detroit, Michigan
22	Elevance Health	Healthcare	156,595	▲ 13.0%	102,200	Indianapolis, Indiana
23	JPMorgan Chase	Financial services	154,792	▲ 21.7%	293,723	New York City, New York
24	Kroger	Retail	148,258	▲ 7.5%	430,000	Cincinnati, Ohio
25	Centene	Healthcare	144,547	▲ 14.7%	74,300	St. Louis, Missouri
26	Verizon Communications	Telecommunications	136,835	▲ 2.4%	117,100	New York City, New York

28	Fannie Mae	Financials	121,596	▲ 19.7%	8,000	Washington, D.C.
29	Comcast	Telecommunications	121,427	▲ 4.3%	186,000	Philadelphia, Pennsylvania
30	AT&T	Conglomerate	120,741	▼ 28.5%	160,700	Dallas, Texas
31	Meta Platforms	Technology	116,609	▼ 1.1%	86,482	Menlo Park, California
32	Bank of America	Financials	115,053	▲ 22.6%	216,823	Charlotte, North Carolina
33	Target Corporation	Retail	109,120	▲ 2.9%	440,000	Minneapolis, Minnesota
34	Dell Technologies	Technology	102,301	▼ 4.4%	133,000	Round Rock, Texas
35	Archer Daniels Midland	Food industry	101,556	▲ 19.1%	41,181	Chicago, Illinois
36	Citigroup	Financials	101,078	▲ 26.6%	238,104	New York City, New York
37	United Parcel Service	Transportation	100,338	▲ 3.1%	404,700	Atlanta, Georgia
38	Pfizer	Pharmaceutical industry	100,330	▲ 23.4%	83,000	New York City, New York
39	Lowe's	Retail	97,059	▲ 0.8%	244,500	Mooresville, North Carolina
40	Johnson & Johnson	Pharmaceutical industry	94,943	▲ 1.2%	152,700	New Brunswick, New Jersey
41	FedEx	Transportation	93,512	▲ 11.4%	518,249	Memphis, Tennessee
42	Humana	Health Insurance	92,870	▲ 11.8%	67,100	Louisville, Kentucky
43	Energy Transfer Partners	Petroleum industry	89,876	▲ 33.3%	12,565	Dallas, Texas
44	State Farm	Financials	89,328	▲ 8.6%	60,519	Bloomington, Illinois
45	Freddie Mac	Financials	86,717	▲ 31.6%	7,819	McLean, Virginia
46	PepsiCo	Beverage	86,859	▲ 8.7%	315,000	Purchase, New York
47	Wells Fargo	Financials	82,859	▲ 0.5%	238,000	San Francisco, California
48	The Walt Disney Company	Media	82,722	▲ 22.7%	195,800	Burbank, California
49	ConocoPhillips	Petroleum industry	82,156	▲ 69.9%	9,500	Houston, Texas
50	Tesla	Automotive and Energy	81,462	▲ 51.4%	127,855	Austin, Texas
51	Procter & Gamble	Consumer products Manufacturing	80,187	▲ 5.3%	106,000	Cincinnati, Ohio
52	United States Postal Service	Logistics	78,620	▲ 2.0%	576,000	Washington, D.C.
53	Albertsons	Retail	77,650	▲ 8.0%	198,650	Boise, Idaho
54	General Electric	Conglomerate	76,555	▲ 3.2%	172,000	Boston, Massachusetts
55	MetLife	Financials	69,898	▼ 1.7%	45,000	New York City, New York
56	Goldman Sachs	Financials	68,711	▲ 5.7%	48,500	New York City, New York

59	RTX Corporation	Conglomerate	67,074	▲ 4.2%	182,000	Arlington County, Virginia
60	Boeing	Aerospace and defense	66,608	▲ 6.9%	156,000	Chicago, Illinois
61	StoneX Group	Financials	66,036	▲ 55.3%	305	New York City, New York
62	Lockheed Martin	Aerospace and Defense	65,984	▼ 1.6%	116,000	Bethesda, Maryland
63	Morgan Stanley	Financials	65,936	▲ 7.9%	82,427	New York City, New York
64	Intel	Technology	63,054	▼ 20.1%	131,900	Santa Clara, California
65	HP	Technology	62,983	▼ 0.8%	58,000	Palo Alto, California
66	TD Synnex	Infotech	62,344	▲ 97.2%	28,500	Clearwater, Florida
67	IBM	Technology	60,530	▼ 16.3%	303,100	Armonk, New York
68	HCA Healthcare	Healthcare	60,233	▲ 2.5%	250,500	Nashville, Tennessee
69	Prudential Financial	Financials	60,050	▼ 15.3%	39,583	Newark, New Jersey
70	Caterpillar	Machinery	59,427	▲ 16.6%	109,100	Deerfield, Illinois
71	Merck & Co.	Pharmaceutical industry	59,283	▲ 15.8%	68,000	Kenilworth, New Jersey
72	World Fuel Services	Petroleum industry and Logistics	59,043	▲ 88.4%	5,214	Miami, Florida
73	New York Life Insurance Company	Insurance	58,445	▲ 14.2%	15,050	New York City, New York
74	Enterprise Products	Petroleum industry	58,186	▲ 42.6%	7,300	Houston, Texas
75	AbbVie	Pharmaceutical industry	58,054	▲ 3.3%	50,000	Lake Bluff, Illinois
76	Plains All American Pipeline	Petroleum industry	57,342	▲ 36.3%	4,100	Houston, Texas



75	AbbVie	Pharmaceutical industry	58,054	▲ 3.3%	50,000	Lake Bluff, Illinois
76	Plains All American Pipeline	Petroleum industry	57,342	▲ 36.3%	4,100	Houston, Texas
77	Dow Chemical Company	Chemical industry	56,902	▲ 3.5%	37,800	Midland, Michigan
78	AIG	Insurance	56,437	▲ 8.4%	26,200	New York City, New York
79	American Express	Financial	55,625	▲ 27.3%	77,300	New York City, New York
80	Publix	Retail	54,942	▲ 13.5%	242,000	Lakeland, Florida
81	Charter Communications	Telecommunications	54,022	▲ 4.5%	101,700	Stamford, Connecticut
82	Tyson Foods	Food Processing	53,282	▲ 13.2%	142,000	Springdale, Arkansas
83	John Deere	Agriculture manufacturing	52,577	▲ 19.4%	82,239	Moline, Illinois
84	Cisco	Telecom Hardware Manufacturing	51,557	▲ 3.5%	83,300	San Jose, California
85	Nationwide Mutual Insurance Company	Financial	51,450	▲ 8.6%	24,791	Columbus, Ohio
86	Allstate	Insurance	51,412	▼ 3.4%	54,250	Northfield Township, Cook County, Illinois
87	Delta Air Lines	Airline	50,582	▲ 69.2%	95,000	Atlanta, Georgia
88	Liberty Mutual	Insurance	49,956	▲ 3.6%	50,000	Boston, Massachusetts
89	TJX	Retail	49,936	▲ 2.9%	329,000	Framingham, Massachusetts
90	Progressive Corporation	Insurance	49,611	▲ 4.0%	55,063	Mayfield Village, Ohio
91	American Airlines	Airline	48,971	▲ 63.9%	129,700	Fort Worth, Texas
92	CHS	Agriculture cooperative	47,194	▲ 24.3%	10,014	Inver Grove Heights, Minnesota
93	Performance Food Group	Food Processing	47,194	▲ 61.6%	34,825	Richmond, Virginia
94	PBF Energy	Petroleum industry	46,830	▲ 71.8%	3,616	Parsippany–Troy Hills, New Jersey
95	Nike	Apparel	46,710	▲ 4.9%	79,100	Beaverton, Oregon
96	Best Buy	Retail	46,298	▼ 10.6%	71,100	Richfield, Minnesota
97	Bristol-Myers Squibb	Pharmaceutical industry	46,159	▼ 0.5%	34,300	New York City, New York
98	United Airlines	Airline	44,955	▲ 82.5%	92,795	Chicago, Illinois
99	Thermo Fisher Scientific	Laboratory instruments	44,915	▲ 14.5%	130,000	Waltham, Massachusetts
100	Qualcomm	Technology	44,200	▲ 31.7%	51,000	San Diego, California

Below are the 100 largest companies by revenue in 2023 (mostly for fiscal year 2022), according to the Fortune 500 list.[1]

ANALYSIS: BASIC DESCRIPTIVE & MATHEMATICAL OR STATISTICAL ANALYSIS

1. DESCRIPTIVE ANALYSIS (MEAN, MEDIAN, MODE, STANDARD DEVIATION, ETC)

S No.	PARTICULARS	RANK	VALUES
1	COUNT	100.000000	100.000000
2	MEAN	50.500000	117566.460000
3	STANDARD DEVIATION	29.011492	100915.112163
4	MIN	1.000000	44200.000000
5	25%	25.750000	57876.000000
6	50%	50.500000	80824.500000
7	75%	75.250000	138763.000000
8	MAX	100.000000	611289.000000

Computation of Descriptive Analysis using Python:

```
In [49]: df.describe(include = "all")
```

Out[49]:

	Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
count	100	100	100	100	100	100	100
unique	100	100	38	99	92	97	76
top	1	Walmart	Financials	47,194	9.4%	50,000	New York City, New York
freq	1	1	11	2	2	2	12

```
In [18]: import pandas as pd

# Loading the dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Da

# Removing commas and converting to numeric
data['Revenue (USD millions)'] = data['Revenue (USD millions)'].str.replace(',',

# Calculating the mean of the revenue column
revenue_mean = data['Revenue (USD millions)'].mean()

# Printing the mean
print(f"Mean Revenue: {revenue_mean:.2f} USD millions")
```

```
Mean Revenue: 117566.46 USD millions
```

```
In [19]: # Summary statistics
summary_stats = data.describe()
print(summary_stats)
```

```
              Rank  Revenue (USD millions)
count   100.000000              100.000000
mean     50.500000           117566.460000
std      29.011492           100915.112163
min       1.000000            44200.000000
25%      25.750000            57876.000000
50%      50.500000            80824.500000
75%      75.250000           138763.000000
max     100.000000           611289.000000
```

- This code snippet calculates summary statistics for the dataset, providing us with essential insights into the central tendency and distribution of our key numerical variables, such as revenue and revenue growth.

- After reviewing these summary statistics, the next steps could include more in-depth analysis, visualization, or data preprocessing based on the specific objectives of the project.

In [63]:
```
import pandas as pd

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Remove commas and convert the column to numeric
data['Revenue (USD millions)'] = data['Revenue (USD millions)'].str.replace(',', '').astype(float)

# Calculate the median of the 'Revenue (USD millions)' column
median_revenue = data['Revenue (USD millions)'].median()

# Print the median
print(f"Median Revenue: {median_revenue:.2f} USD millions")
```

```
Median Revenue: 80824.50 USD millions
```

In [64]:
```
import pandas as pd

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Remove commas and convert the 'Revenue (USD millions)' column to numeric
data['Revenue (USD millions)'] = data['Revenue (USD millions)'].str.replace(',', '').astype(float)

# Calculate the variance of the 'Revenue (USD millions)' column
revenue_variance = data['Revenue (USD millions)'].var()

# Print the variance
print(f"Variance of Revenue: {revenue_variance:.2f} USD millions^2")
```

```
Variance of Revenue: 10183859862.96 USD millions^2
```

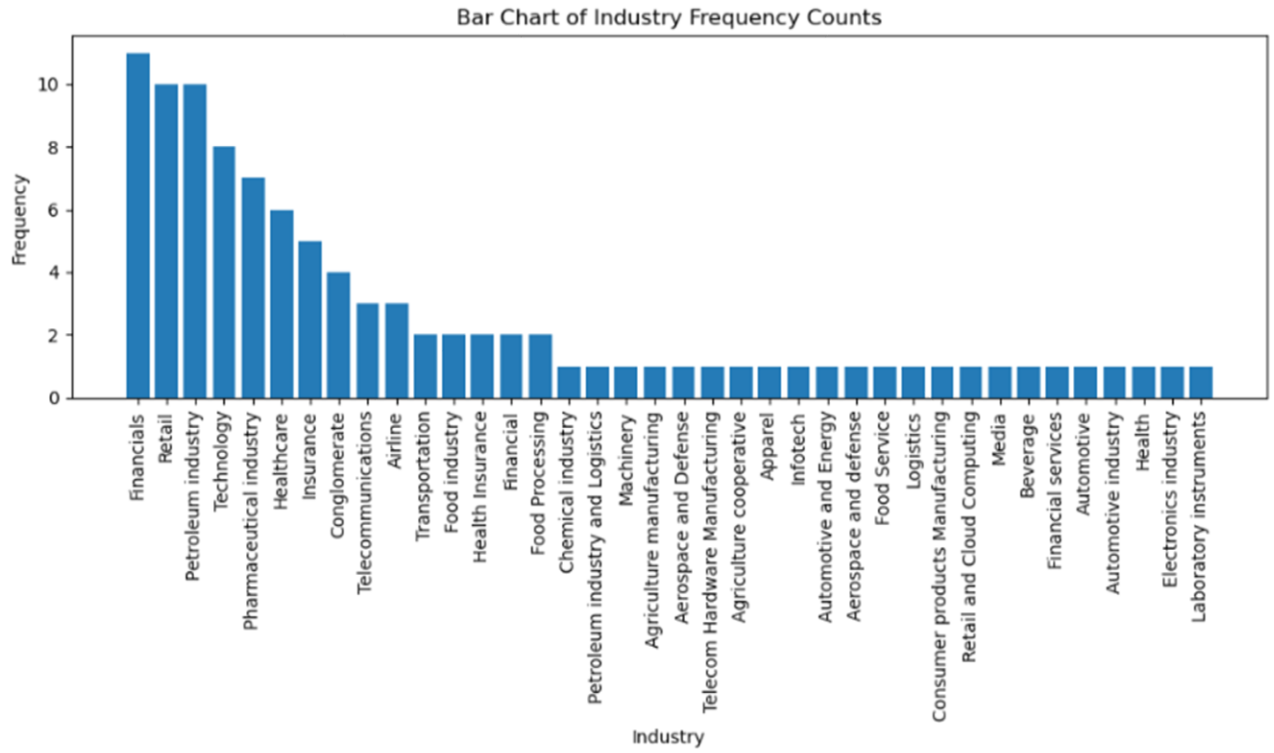
In [66]:
```
import pandas as pd

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Remove commas and convert the 'Revenue (USD millions)' column to numeric
data['Revenue (USD millions)'] = data['Revenue (USD millions)'].str.replace(',', '').astype(float)

# Calculate the standard deviation of the 'Revenue (USD millions)' column
revenue_std_dev = data['Revenue (USD millions)'].std()

# Print the standard deviation
print(f"Standard Deviation of Revenue: {revenue_std_dev:.2f} USD millions")
```

```
Standard Deviation of Revenue: 100915.11 USD millions
```

- As an extension to the previous analysis, there 2 snippets contain the following:
  - Median Revenue – 80824.50 USD millions
  - Variance of Revenue – 10183859862.96 USD millions
  - Standard Deviation of Revenue – 100915.11 USD millions.

2. DATA VISUALIZATION (CHARTS)

In [22]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Frequency counts for the 'Industry' categorical column
industry_counts = data['Industry'].value_counts()

# Create a bar chart for the frequency counts
plt.figure(figsize=(10, 6))  # Optional: Adjust the figure size
plt.bar(industry_counts.index, industry_counts)
plt.xlabel('Industry')
plt.ylabel('Frequency')
plt.title('Bar Chart of Industry Frequency Counts')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability (optional)
plt.tight_layout()  # Optional: Adjust spacing for labels
plt.show()
```

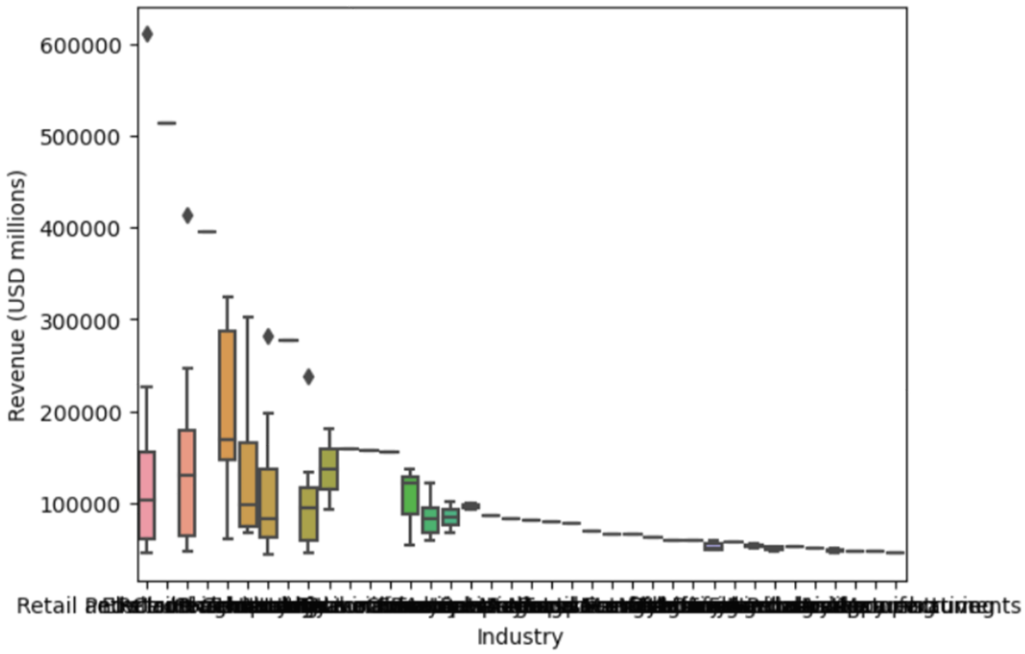


This code provided a bar chart displaying the frequency counts of different industries in the dataset. Based on this output, here are some potential managerial implications:

- ✓ This chart provides a clear visual representation of how frequently each industry appears in the dataset. This can be valuable for managers to understand which industries are most prevalent among the largest companies in the United States.
- ✓ Managers can use this information to assess whether the company's current market focus aligns with the industries represented in the dataset.
- ✓ If the company is heavily concentrated in one industry and there are other industries with high representation in the dataset, it may suggest opportunities for diversification. Diversification can spread risk and potentially lead to new revenue streams.
- ✓ Understanding which industries are dominant among the largest companies allows managers to benchmark their company's performance against competitors in those industries.

- ✓ For companies operating in multiple industries, this information can inform marketing and sales strategies. Tailoring approaches to specific industries can lead to more effective customer acquisition and retention.
- ✓ Industries with high representation might present opportunities for strategic alliances or partnerships. Collaborating with key players in these industries can lead to mutually beneficial outcomes. Industries with low representation may represent a gap in the company's current portfolio.
- ✓ Managers should assess whether this gap presents an opportunity or if it exposes the company to unnecessary risk.

```
In [39]: import seaborn as sns
         # Convert the 'Revenue (USD millions)' column to numeric
         data['Revenue (USD millions)'] = pd.to_numeric(data['Revenue (USD millions)'], errors='coerce')
         # Create the boxplot
         sns.boxplot(x='Industry', y='Revenue (USD millions)', data=data)

Out[39]: <Axes: xlabel='Industry', ylabel='Revenue (USD millions)'>
```



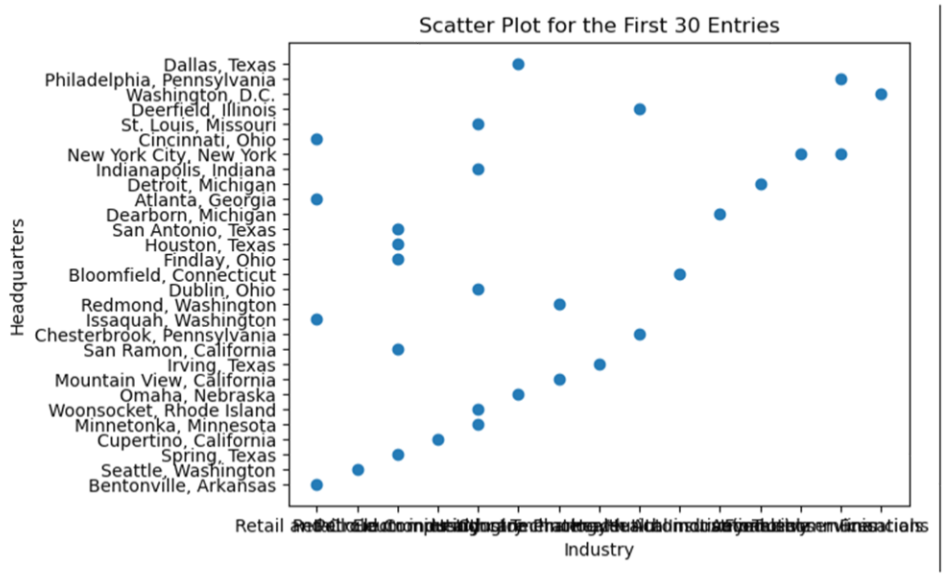
The code provided a boxplot to visualize the distribution of revenue (in USD millions) across different industries. Maximum revenue was earned by Walmart in the Retail Industry i.e., $611289 mn.

Here are the managerial implications of the output from this chart:

- ✓ The boxplot helps managers understand the variability in revenue across different industries. It provides a visual representation of the spread of revenue values within each industry category.
- ✓ Outliers in the boxplot can be indicative of companies that have exceptionally high or low revenues compared to their industry peers. Managers should investigate these outliers to understand the factors contributing to their unique performance.
- ✓ The distribution of revenue within industries can guide strategic decisions related to resource allocation, market expansion, and investment. It helps identify industries with growth potential and those that may require special attention.

- ✓ The provided code generates a count plot of the 'Industry' distribution for the first 30 entries in your dataset. This plot visualizes the frequency of each industry category within this subset of data. The managerial implications of this visualization can include:

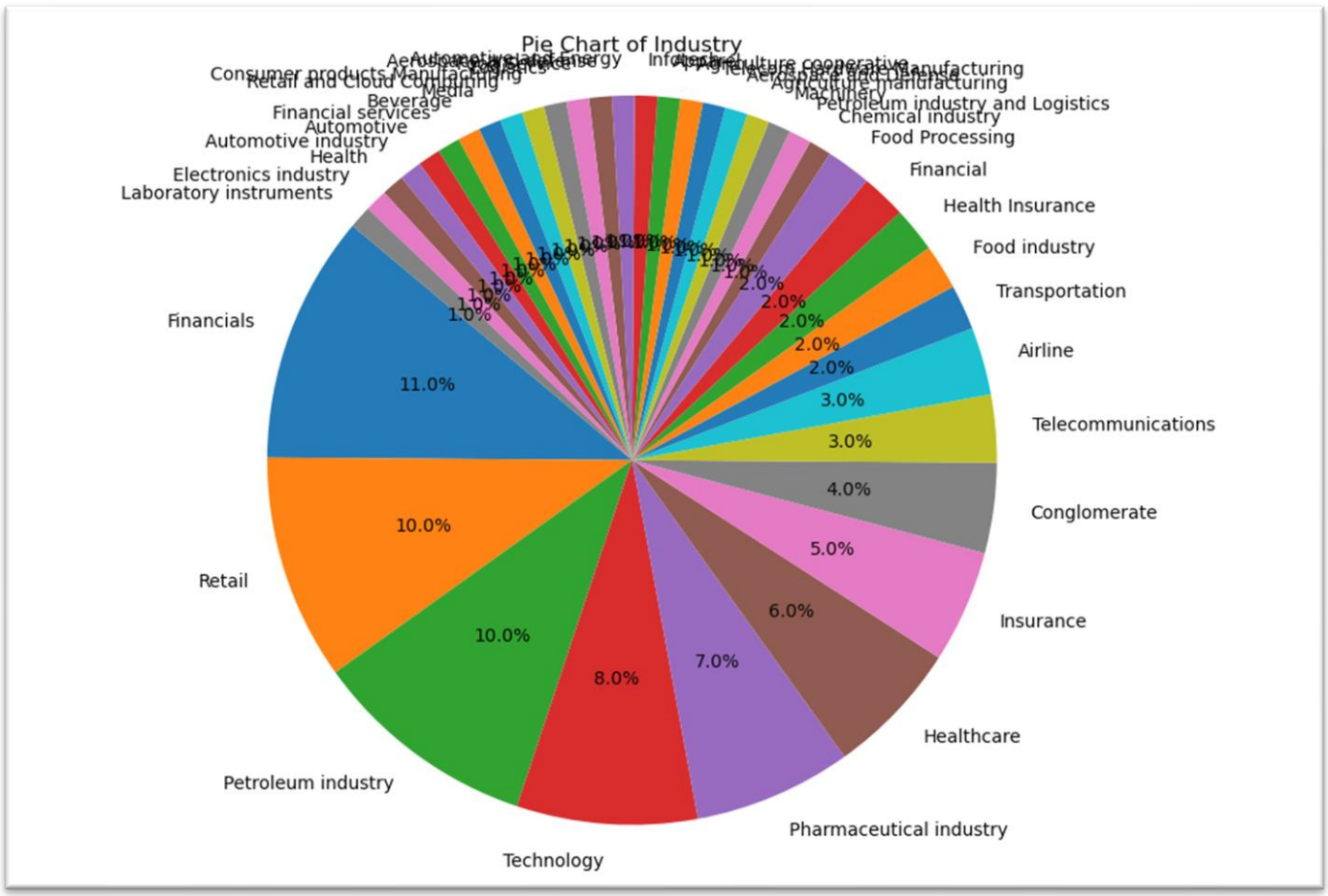
```
In [53]: import matplotlib.pyplot as plt

         # Select the first 30 entries in your dataset
         subset_data = data.head(30)

         # Scatter Plot
         plt.scatter(subset_data['Industry'], subset_data['Headquarters'])
         plt.xlabel('Industry')
         plt.ylabel('Headquarters')

         plt.title('Scatter Plot for the First 30 Entries')
         plt.show()
         print(data.columns)
```




```
Index(['Rank', 'Name', 'Industry', 'Revenue (USD millions)', 'Revenue growth',
       'Employees', 'Headquarters'],
      dtype='object')
```

- ✓ The scatter plot helps identify which industries have a significant presence in the first 30 entries. Managers can focus their attention on understanding and addressing the needs and challenges specific to these dominant industries.
- ✓ By observing the distribution of industries, managers can gain insights into potential market segments or customer groups that are more prevalent in the early portion of the dataset. This information can guide marketing and sales strategies.
- ✓ For companies operating in multiple industries, this visualization can help managers assess the diversity of their business portfolio. It may trigger discussions about resource allocation, risk management, or diversification strategies.

```
In [56]: import pandas as pd
         import matplotlib.pyplot as plt

         # Load your dataset
         data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

         # Calculate category counts
         category_counts = data['Industry'].value_counts()

         # Create a pie chart
         plt.figure(figsize=(8, 8))  # Optional: Adjust the figure size
         plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', startangle=140)
         plt.title('Pie Chart of Industry')
         plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.

         # Display the chart
         plt.show()
```

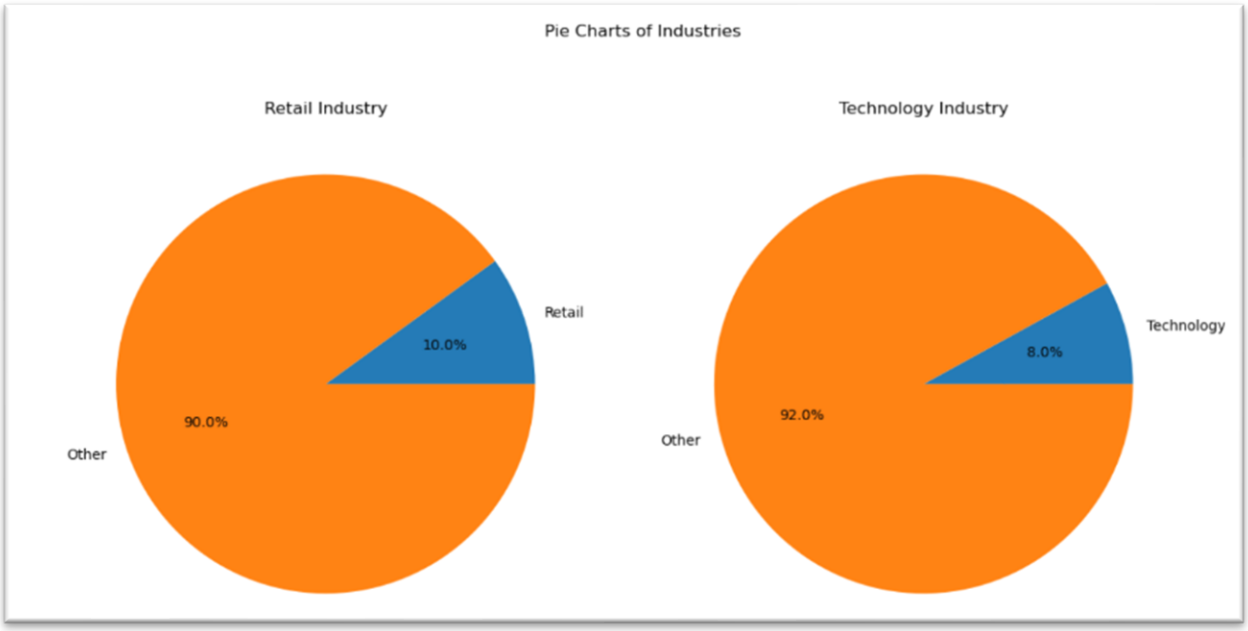


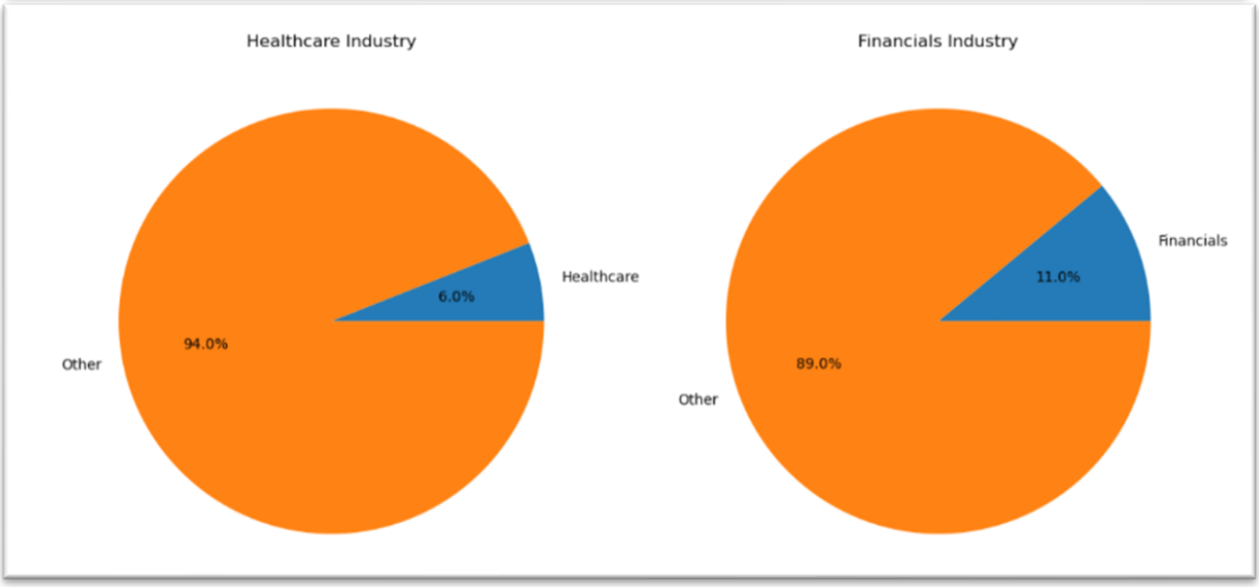
INDUSTRY	PERCENTAGE
FINANCIALS	11.0%
RETAIL	10.0%
PETROLEUM INDUSTRY	10.0%
TECHNOLOGY	8.0%
PHARMACEUTICAL INDUSTRY	7.0%
HEALTHCARE	6.0%
INSURANCE	5.0%
CONGLOMERATE	4.0%
TELECOMMUNICATIONS	3.0%

AIRLINE	3.0%
TRANSPORTATION, FOOD INDUSTRY, HEALTH INSURANCE, FINANCIAL, FOOD PROCESSING	2.0%
REMAINING 23 INDUSTRIES	1.0%

- ✓ The pie chart created displays the distribution of companies across different industries based on the dataset. The managerial implications of this chart would depend on the context and the goals of the analysis, but here are some general insights and implications that can be drawn:
- ✓ The chart provides a visual representation of how companies are distributed among various industries. Managers can use this information to gain a better understanding of the composition of industries in the dataset.
- ✓ It's clear from the chart which industries have a larger presence in the dataset. This information can be valuable for decision-making, such as when considering industry-specific trends or competition.
- ✓ Managers can assess the level of concentration in certain industries. For example, if a single industry dominates the chart, it may indicate a highly concentrated market, which could have implications for competitive dynamics.
- ✓ For companies looking to diversify their operations, the chart can highlight industries with lower representation. Exploring opportunities in these less-represented industries may be a strategic consideration.
- ✓ This chart can be useful in long-term planning, helping companies align their goals with the industry landscape and anticipate potential shifts in the market.
- ✓ Managers should interpret the chart in the context of their organization's goals and strategies.

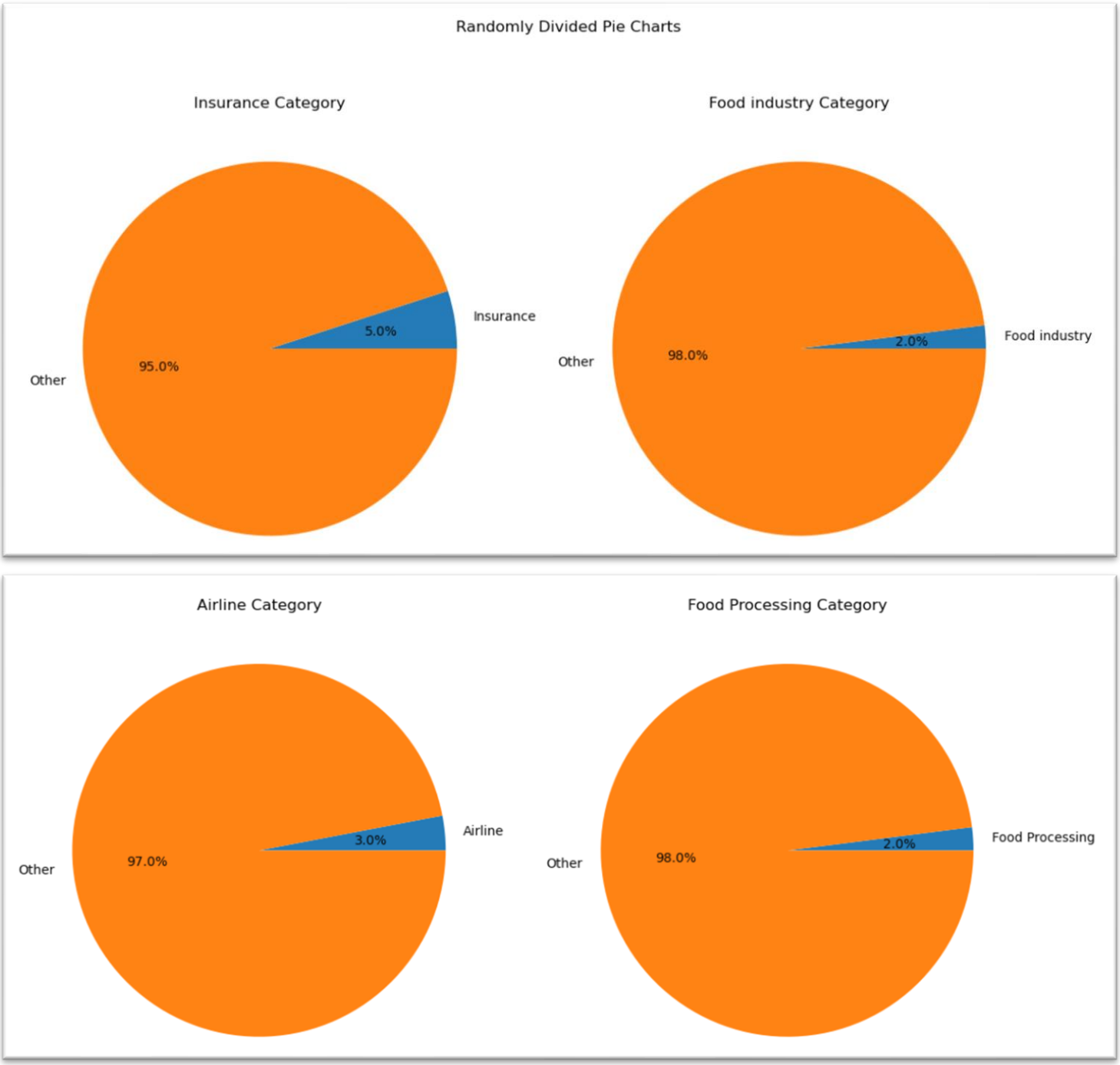
In [59]:
```
import matplotlib.pyplot as plt
import pandas as pd

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Define the industries you want to create pie charts for (customize as needed)
industries = ['Retail', 'Technology', 'Healthcare', 'Financials']

# Create subplots for the pie charts
fig, axs = plt.subplots(2, 2, figsize=(12, 12))
fig.suptitle('Pie Charts of Industries')

# Iterate through the selected industries and create pie charts
for i, industry in enumerate(industries):
    # Filter data for the current industry
    industry_data = data[data['Industry'] == industry]

    # Calculate the counts of companies in the current industry
    industry_counts = len(industry_data)

    # Plot the pie chart in the corresponding subplot
    row = i // 2
    col = i % 2
    ax = axs[row, col]
    ax.pie([industry_counts, len(data) - industry_counts], labels=[industry, 'Other'], autopct='%1.1f%%')
    ax.set_title(f'{industry} Industry')

# Adjust spacing between subplots
plt.tight_layout()
plt.subplots_adjust(top=0.9)

# Show the pie charts
plt.show()
```





In [61]:

```python
import matplotlib.pyplot as plt
import pandas as pd
import random

# Load your dataset
data = pd.read_csv('List_of_largest_companies_in_the_United_States_by_revenue_Data.csv')

# Get a list of unique industry categories
unique_categories = data['Industry'].unique()

# Define the four categories randomly (customize as needed)
categories = random.sample(list(unique_categories), 4)

# Create subplots for the pie charts
fig, axs = plt.subplots(2, 2, figsize=(12, 12))
fig.suptitle('Randomly Divided Pie Charts')

# Iterate through the selected categories and create pie charts
for i, category in enumerate(categories):
    # Filter data for the current category
    category_data = data[data['Industry'] == category]

    # Calculate the counts of companies in the current category
    category_counts = len(category_data)

    # Plot the pie chart in the corresponding subplot
    row = i // 2
    col = i % 2
    ax = axs[row, col]
    ax.pie([category_counts, len(data) - category_counts], labels=[category, 'Other'], autopct='%1.1f%%')
    ax.set_title(f'{category} Category')

# Adjust spacing between subplots
plt.tight_layout()
plt.subplots_adjust(top=0.9)

# Show the pie charts
plt.show()
```



✓ Also, 4 specific and 4 random pie-charts are made out of all the industries.

FINDINGS & INFERENCES

From the provided dataset of the largest companies in the United States by revenue, here are some findings and inferences:

1. **INDUSTRY DISTRIBUTION:**
   - The dataset covers a diverse range of industries, including Retail, Technology, Petroleum, Healthcare, Financials, and more.
   - Retail appears to be a dominant industry among the largest companies.

List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington

2. **REVENUE RANKING:**
   - Walmart tops the list with the highest revenue, followed by Amazon, Exxon Mobil, and Apple.
   - Companies in the Retail and Technology sectors are prominent in the top rankings.

List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington

3. REVENUE GROWTH:

- Some companies, like Exxon Mobil, Chevron Corporation, and ConocoPhillips in the Petroleum industry, have seen significant revenue growth.
- Technology companies like Meta Platforms (formerly Facebook) and Tesla also show substantial growth.

List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington

30	AT&T	Conglomerate	120,741	▼ 28.5%	160,700	Dallas, Texas
31	Meta Platforms	Technology	116,609	▼ 1.1%	86,482	Menlo Park, California
32	Bank of America	Financials	115,053	▲ 22.6%	216,823	Charlotte, North Carolina
33	Target Corporation	Retail	109,120	▲ 2.9%	440,000	Minneapolis, Minnesota
34	Dell Technologies	Technology	102,301	▼ 4.4%	133,000	Round Rock, Texas
35	Archer Daniels Midland	Food industry	101,556	▲ 19.1%	41,181	Chicago, Illinois
36	Citigroup	Financials	101,078	▲ 26.6%	238,104	New York City, New York
37	United Parcel Service	Transportation	100,338	▲ 3.1%	404,700	Atlanta, Georgia
38	Pfizer	Pharmaceutical industry	100,330	▲ 23.4%	83,000	New York City, New York
39	Lowe's	Retail	97,059	▲ 0.8%	244,500	Mooresville, North Carolina
40	Johnson & Johnson	Pharmaceutical industry	94,943	▲ 1.2%	152,700	New Brunswick, New Jersey
41	FedEx	Transportation	93,512	▲ 11.4%	518,249	Memphis, Tennessee
42	Humana	Health Insurance	92,870	▲ 11.8%	67,100	Louisville, Kentucky
43	Energy Transfer Partners	Petroleum industry	89,876	▲ 33.3%	12,565	Dallas, Texas
44	State Farm	Financials	89,328	▲ 8.6%	60,519	Bloomington, Illinois
45	Freddie Mac	Financials	86,717	▲ 31.6%	7,819	McLean, Virginia
46	PepsiCo	Beverage	86,859	▲ 8.7%	315,000	Purchase, New York
47	Wells Fargo	Financials	82,859	▲ 0.5%	238,000	San Francisco, California
48	The Walt Disney Company	Media	82,722	▲ 22.7%	195,800	Burbank, California
49	ConocoPhillips	Petroleum industry	82,156	▲ 69.9%	9,500	Houston, Texas
50	Tesla	Automotive and Energy	81,462	▲ 51.4%	127,855	Austin, Texas
51	Procter & Gamble	Consumer products Manufacturing	80,187	▲ 5.3%	106,000	Cincinnati, Ohio

4. EMPLOYEE COUNT:

- Walmart and Amazon employ a massive number of people, with over a million employees each.
- Technology companies like Apple and Microsoft have comparatively fewer employees but still play a significant role in the economy.

List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington
13	Microsoft	Technology	198,270	▲ 18.0%	221,000	Redmond, Washington
14	Cardinal Health	Healthcare	181,364	▲ 11.6%	46,035	Dublin, Ohio
15	Cigna	Health Insurance	180,516	▲ 3.7%	70,231	Bloomfield, Connecticut

5. GEOGRAPHIC DISTRIBUTION:

- Headquarters of these large companies are spread across the United States, including cities like Seattle, San Francisco, New York City, and Houston.

2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
32	Bank of America	Financials	115,053	▲ 22.6%	216,823	Charlotte, North Carolina
33	Target Corporation	Retail	109,120	▲ 2.9%	440,000	Minneapolis, Minnesota
34	Dell Technologies	Technology	102,301	▼ 4.4%	133,000	Round Rock, Texas
35	Archer Daniels Midland	Food industry	101,556	▲ 19.1%	41,181	Chicago, Illinois
36	Citigroup	Financials	101,078	▲ 26.6%	238,104	New York City, New York
37	United Parcel Service	Transportation	100,338	▲ 3.1%	404,700	Atlanta, Georgia
38	Pfizer	Pharmaceutical industry	100,330	▲ 23.4%	83,000	New York City, New York
39	Lowe's	Retail	97,059	▲ 0.8%	244,500	Mooresville, North Carolina
40	Johnson & Johnson	Pharmaceutical industry	94,943	▲ 1.2%	152,700	New Brunswick, New Jersey
41	FedEx	Transportation	93,512	▲ 11.4%	518,249	Memphis, Tennessee
42	Humana	Health Insurance	92,870	▲ 11.8%	67,100	Louisville, Kentucky
43	Energy Transfer Partners	Petroleum industry	89,876	▲ 33.3%	12,565	Dallas, Texas
44	State Farm	Financials	89,328	▲ 8.6%	60,519	Bloomington, Illinois
45	Freddie Mac	Financials	86,717	▲ 31.6%	7,819	McLean, Virginia
46	PepsiCo	Beverage	86,859	▲ 8.7%	315,000	Purchase, New York
47	Wells Fargo	Financials	82,859	▲ 0.5%	238,000	San Francisco, California
48	The Walt Disney Company	Media	82,722	▲ 22.7%	195,800	Burbank, California
49	ConocoPhillips	Petroleum industry	82,156	▲ 69.9%	9,500	Houston, Texas
50	Tesla	Automotive and Energy	81,462	▲ 51.4%	127,855	Austin, Texas

6. **HEALTHCARE & PHARMACEUTICAL:**
   - Healthcare companies like UnitedHealth Group and CVS Health feature prominently in the top rankings.
   - AmerisourceBergen and Pfizer represent the pharmaceutical industry.

List of the largest companies [edit]

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
1	Walmart	Retail	611,289	▲ 6.7%	2,100,000	Bentonville, Arkansas
2	Amazon	Retail and Cloud Computing	513,983	▲ 9.4%	1,540,000	Seattle, Washington
3	Exxon Mobil	Petroleum industry	413,680	▲ 44.8%	62,000	Spring, Texas
4	Apple	Electronics industry	394,328	▲ 7.8%	164,000	Cupertino, California
5	UnitedHealth Group	Healthcare	324,162	▲ 12.7%	400,000	Minnetonka, Minnesota
6	CVS Health	Healthcare	322,467	▲ 10.4%	259,500	Woonsocket, Rhode Island
7	Berkshire Hathaway	Conglomerate	302,089	▲ 9.4%	383,000	Omaha, Nebraska
8	Alphabet	Technology	282,836	▲ 9.8%	156,000	Mountain View, California
9	McKesson Corporation	Health	276,711	▲ 4.8%	48,500	Irving, Texas
10	Chevron Corporation	Petroleum industry	246,252	▲ 51.6%	43,846	San Ramon, California
11	AmerisourceBergen	Pharmaceutical industry	238,587	▲ 11.5%	41,500	Chesterbrook, Pennsylvania
12	Costco	Retail	226,954	▲ 15.8%	304,000	Issaquah, Washington
38	Pfizer	Pharmaceutical industry	100,330	▲ 23.4%	83,000	New York City, New York

7. **FINANCIAL SECTOR:**
   - Several companies in the financial sector, such as JPMorgan Chase, Bank of America, and Citigroup, are among the top 50.

23	JPMorgan Chase	Financial services	154,792	▲ 21.7%	293,723	New York City, New York
32	Bank of America	Financials	115,053	▲ 22.6%	216,823	Charlotte, North Carolina
36	Citigroup	Financials	101,078	▲ 26.6%	238,104	New York City, New York

❖ These findings provide insights into the composition of the largest U.S. companies by revenue, their respective industries, and their performance in terms of revenue and employee count. These companies collectively play a crucial role in the U.S. economy and various sectors.

MANAGERIAL INSIGHTS | IMPLICATIONS

Here are some overall managerial insights and implications from the dataset of the largest companies in the United States by revenue:

1. **INDUSTRY TRENDS:**
   ✓ Understanding the dominant industries among the largest companies can help business leaders identify lucrative sectors and potential areas for investment.

2. **REVENUE GROWTH:**
   - ✓ Analysing revenue growth rates can provide insights into industry performance.
   - ✓ Rapidly growing industries may present expansion opportunities.

3. **COMPETITIVE ANALYSIS:**
   - ✓ Identifying top companies in each industry allows for benchmarking and understanding competitive dynamics within sectors.

4. **LABOR MARKET INSIGHTS:**
   - ✓ Companies with extensive employee bases can influence labor markets.
   - ✓ Monitoring these companies can help in assessing workforce trends and labor competition.

5. **INVESTMENT DECISIONS:**
   - ✓ Investors can use this data to inform their investment decisions, focusing on industries and companies with strong revenue growth and stability.

6. **STRATEGIC PARTNERSHIPS:**
   - ✓ Companies looking to form strategic partnerships or collaborations can target those in related industries for mutually beneficial ventures.

7. **MERGERS AND ACQUISITIONS:**
   - ✓ Businesses considering mergers or acquisitions can identify potential targets based on their industry and revenue rankings.

8. **SUPPLY CHAIN MANAGEMENT:**
   - ✓ Understanding the major players in industries like Petroleum, Technology, and Retail can inform supply chain decisions and vendor relationships.

9. **MARKET EXPANSION:**
   - ✓ Companies seeking to expand their market presence can evaluate regions where these large companies are headquartered.

10. **SECTOR RESILIENCE:**
    - ✓ The dataset provides insights into how different industries, such as Healthcare and Technology, have fared during various economic conditions.
