

List of companies in the United States of America

```
In [1]: from bs4 import BeautifulSoup
import requests

In [24]: url = "https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue"

page = requests.get(url)

soup = BeautifulSoup(page.text, 'html')

print(soup)

In [25]: # find the table of interest
soup.find_all('table')[1]
#<table class="wikitable sortable jquery-tablesorter">

In [10]: # Give the table a name.. in this case it is called table
table = soup.find_all('table')[1]

In [27]: #find column headers from the table
Table_headers = table.find_all('th')
Table_headers

In [13]: #Clean the column header details by removing <th>
Table_column_headers = [title.text.strip() for title in Table_headers]
print(Table_column_headers)

['Rank', 'Name', 'Industry', 'Revenue (USD millions)', 'Revenue growth', 'Employees', 'Headquarters']

In [15]: # Import the column headers to pandas
import pandas as pd

In [18]: #Create a dataframe and insert the table colmn headers into it
df = pd.DataFrame(columns = Table_column_headers)
df
```

Out[18]:

Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
------	------	----------	------------------------	----------------	-----------	--------------

```
In [26]: #Finding columns data
Column_data = table.find_all('tr')
Column_data

In [22]: #Column_data_all =
for row in Column_data[1:]:
    row_data = row.find_all('td')
    individual_row_data = [data.text.strip() for data in row_data]

    length = len(df)
    df.loc[length] = individual_row_data
```

```
In [23]: df
```

Out[23]:

	Rank	Name	Industry	Revenue (USD millions)	Revenue growth	Employees	Headquarters
0	1	Walmart	Retail	611,289	6.7%	2,100,000	Bentonville, Arkansas
1	2	Amazon	Retail and cloud computing	513,983	9.4%	1,540,000	Seattle, Washington
2	3	ExxonMobil	Petroleum industry	413,680	44.8%	62,000	Spring, Texas
3	4	Apple	Electronics industry	394,328	7.8%	164,000	Cupertino, California
4	5	UnitedHealth Group	Healthcare	324,162	12.7%	400,000	Minnetonka, Minnesota
...
95	96	Best Buy	Retail	46,298	10.6%	71,100	Richfield, Minnesota
96	97	Bristol-Myers Squibb	Pharmaceutical industry	46,159	0.5%	34,300	New York City, New York
97	98	United Airlines	Airline	44,955	82.5%	92,795	Chicago, Illinois
98	99	Thermo Fisher Scientific	Laboratory instruments	44,915	14.5%	130,000	Waltham, Massachusetts
99	100	Qualcomm	Technology	44,200	31.7%	51,000	San Diego, California

100 rows × 7 columns