```
In [2]:    1  from bs4 import BeautifulSoup
           2  import requests
```

```
In [3]:    1  url='https://en.wikipedia.org/wiki/List_of_largest_companies_in_India'
           2  page=requests.get(url)
           3  soup=BeautifulSoup(page.text,'html')
```

```
In [4]:    1  print(soup)
```

```
<!DOCTYPE html>
<html class="client-nojs vector-feature-language-in-header-enabled vector-feature-language-in-main-page-he
ader-disabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-disabled vector-featu
re-toc-pinned-enabled vector-feature-main-menu-pinned-disabled vector-feature-limited-width-clientpref-1 v
ector-feature-limited-width-content-enabled vector-feature-zebra-design-disabled" dir="ltr" lang="en">
<head>
<meta charset="utf-8"/>
<title>List of largest companies in India - Wikipedia</title>
<script>(function(){var className="client-js vector-feature-language-in-header-enabled vector-feature-lang
uage-in-main-page-header-disabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-d
isabled vector-feature-toc-pinned-enabled vector-feature-main-menu-pinned-disabled vector-feature-limited-
width-clientpref-1 vector-feature-limited-width-content-enabled vector-feature-zebra-design-disabled";var
cookie=document.cookie.match(/(?:^|; )enwikimwclientpreferences=([^;]+)/);if(cookie){cookie[1].split('%2C
').forEach(function(pref){className=className.replace(new RegExp('(^| )'+pref.replace(/-clientpref-\w+$|[^
\w-]+/g,'')+'-clientpref-\\w+( |$)'),'$1'+pref+'$2');});}document.documentElement.className=classNam
e;}());RLCONF={"wgBreakFrames":false,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"
wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February","March","April","May","June","July","Au
gust","September","October","November","December"],
"wgRequestId":"6be3c21f-9a99-4345-a50d-9e5c4bb78283","wgCanonicalNamespace":"","wgCanonicalSpecialPageNam
```

```
In [6]:   1  soup.find_all('table')[1]
```

Out[6]:
```html
<table class="wikitable sortable">
<tbody><tr>
<th>Rank
</th>
<th>Name
</th>
<th>Industry
</th>
<th>Revenue<br/><small>(in  <b>₹ Crore</b>)</small>
</th>
<th>Revenue growth
</th>
<th>Profits<br/><small>(in  <b>₹ Crore</b>)</small>
</th>
<th>Headquarters
</th>
<th><a href="/wiki/List_of_public_sector_undertakings_in_India" title="List of public sector undertakings
in India">State Controlled</a>
</th></tr>
```

In [7]:
```python
soup.find('table',class_='wikitable sortable')
```

Out[7]:
```
<table class="wikitable sortable" style="text-align:right;">
<tbody><tr>
<th align="center" colspan="2">Rank
</th>
<th align="center" colspan="2">Forbes <br/>2000 rank
</th>
<th align="center">Name
</th>
<th align="center">Headquarters
</th>
<th align="center">Revenue<br/>(billions <br/>US$)
</th>
<th align="center">Profit<br/>(billions <br/>US$)
</th>
<th align="center">Assets<br/>(billions <br/>US$)
</th>
<th align="center">Value<br/>(billions <br/>US$)
</th>
<th align="center">Industry
```

In [8]:
```python
table=soup.find_all('table')[1]
```

In [10]:
```python
1  print(table)
```

```
<table class="wikitable sortable">
<tbody><tr>
<th>Rank
</th>
<th>Name
</th>
<th>Industry
</th>
<th>Revenue<br/><small>(in  <b>₹ Crore</b>)</small>
</th>
<th>Revenue growth
</th>
<th>Profits<br/><small>(in  <b>₹ Crore</b>)</small>
</th>
<th>Headquarters
</th>
<th><a href="/wiki/List_of_public_sector_undertakings_in_India" title="List of public sector undertakings
in India">State Controlled</a>
</th></tr>
```

In [11]:
```python
1  head=table.find_all('th')
```

In [12]:
```
1  head
```

Out[12]: [<th>Rank
</th>,
<th>Name
</th>,
<th>Industry
</th>,
<th>Revenue
<small>(in ₹ Crore)</small>
</th>,
<th>Revenue growth
</th>,
<th>Profits
<small>(in ₹ Crore)</small>
</th>,
<th>Headquarters
</th>,
<th><a href="/wiki/List_of_public_sector_undertakings_in_India" title="List of public sector undertakings
in India">State Controlled
</th>]

In [14]:
```
1  head1=[title.text for title in head]
2  print(head1)
```

['Rank\n', 'Name\n', 'Industry\n', 'Revenue(in ₹ Crore)\n', 'Revenue growth\n', 'Profits(in ₹ Crore)\n',
'Headquarters\n', 'State Controlled\n']

In [16]:
```
1  head1=[title.text.strip()for title in head]
2  print(head1)
```

['Rank', 'Name', 'Industry', 'Revenue(in ₹ Crore)', 'Revenue growth', 'Profits(in ₹ Crore)', 'Headquarte
rs', 'State Controlled']

In [17]:
```
1  import pandas as pd
```

In [18]:
```python
1  df=pd.DataFrame(columns=head1)
2  df
```

Out[18]:

Rank	Name	Industry	Revenue(in ₹ Crore)	Revenue growth	Profits(in ₹ Crore)	Headquarters	State Controlled

In [19]:
```python
1  column_data=table.find_all('tr')
```

In [24]:
```python
1  for row in column_data[1:]:
2      row_data=row.find_all('td')
3      ind_row_data=[data.text.strip() for data in row_data]
4      #print(ind_row_data)
5      length=len(df)
6      df.loc[length]=ind_row_data
```

In [25]: 1 df

Out[25]:

	Rank	Name	Industry	Revenue(in ₹ Crore)	Revenue growth	Profits(in ₹ Crore)	Headquarters	State Controlled
0	1	Indian Oil Corporation	Oil and gas	424,321	13.2%	22,189	New Delhi	Yes
1	2	Reliance Industries Limited	Oil and gas	410,295	28.2%	36,075	Mumbai	
2	3	Oil and Natural Gas Corporation	Oil and gas	333,143	11.0%	22,106	New Delhi	Yes
3	4	State Bank of India	Banking	306,528	2.6%	−4,556	Mumbai	Yes
4	5	Tata Motors	Automotive	301,175	7.9%	8,989	Mumbai	
5	6	Bharat Petroleum	Oil and gas	238,638	13.7%	9,009	Mumbai	Yes
6	7	Hindustan Petroleum	Oil and gas	221,693	13.4%	7,218	Mumbai	Yes
7	8	Rajesh Exports	Gems and jewellery	187,748	22.5%	1,266	Bangalore	
8	9	Tata Steel	Iron and steel	147,192	25.3%	13,434	Mumbai	
9	10	Coal India	Metals and mining	132,897	5.3%	7,020	Kolkata	Yes
10	11	Tata Consultancy Services	Infotech	126,745	3.6%	25,826	Mumbai	
11	12	Larsen & Toubro	Capital goods	122,743	10.6%	7,370	Mumbai	
12	13	Hindalco Industries	Metals and mining	120,428	15.5%	6,083	Mumbai	
13	14	ICICI Bank	Banking	118,969	4.9%	7,712	Mumbai	
14	15	Vedanta Resources	Metals and mining	102,192	30.8%	10,342	London, UK	
15	16	HDFC Bank	Banking	101,344	17.6%	18,510	Mumbai	
16	17	Mahindra & Mahindra	Automotive	96,377	12.2%	7,510	Mumbai	
17	18	NTPC Limited	Utilities	90,144	7.7%	10,544	New Delhi	Yes
18	19	Bharti Airtel	Telecommunication	86,704	13.0%	1,099	New Delhi	
19	20	Maruti Suzuki India	Automotive	81,978	15.6%	7,880	New Delhi	
20	21	Infosys	Infotech	73,762	3.1%	16,029	Bangalore	
21	22	Nayara Energy	Oil and gas	73,015	10.7%	576	Mumbai	

	Rank	Name	Industry	Revenue(in ₹ Crore)	Revenue growth	Profits(in ₹ Crore)	Headquarters	State Controlled
22	23	HDFC	Financials	72,596	18.9%	16,255	Mumbai	
23	24	JSW Steel Ltd	Iron and steel	70,148	22.6%	6,214	Mumbai	
24	25	Axis Bank	Banking	58,477	1.5%	456	Mumbai	Yes
25	26	Steel Authority of India	Iron and steel	57,626	28.8%	−281	New Delhi	Yes
26	27	Punjab National Bank	Banking	57,608	0.7%	−12,130	New Delhi	Yes
27	28	Grasim Industries	Diversified	57,273	54.9%	2,679	Mumbai	
28	29	Wipro	Infotech	56,986	1.6%	8,003	Bangalore	
29	30	Motherson Sumi Systems	Automotive	56,839	32.7%	1,597	Noida	
30	31	GAIL	Oil and gas	55,503	11.6%	4,799	New Delhi	Yes
31	32	Bank of Baroda	Banking	54,048	3.1%	−1,887	Vadodara	Yes
32	33	HCL Technologies	Infotech	51,695	6.3%	8,721	Noida	
33	34	MRPL	Oil and gas	50,209	8.9%	1,993	Mangalore	Yes
34	35	Canara Bank	Banking	48,743	1.2%	−3,951	Bangalore	Yes
35	36	ITC Limited	Consumer goods	44,674	1.6%	11,271	Kolkata	
36	37	Bank of India	Banking	44,158	4.8%	−5,961	Mumbai	Yes
37	38	General Insurance Corporation of India	Insurance	43,555	38.9%	3,146	Mumbai	Yes
38	39	Kotak Mahindra Bank	Banking	38,813	14.2%	6,201	Mumbai	
39	40	Union Bank of India	Banking	38,413	0.4%	−5,212	Mumbai	Yes
40	41	Hindustan Unilever	Consumer goods	36,088	7.1%	5,214	Mumbai	
41	42	Vodafone India	Telecommunication	34,855	18.9%	7,766	Mumbai	
42	43	Tata Power	Utilities	33,213	12.8%	2,477	Mumbai	
43	44	Chennai Petroleum Corporation	Oil and gas	33,187	20.0%	927	Chennai	Yes
44	45	Hero MotoCorp	Automotive	33,141	12.9%	3,720	New Delhi	
45	46	Tech Mahindra	Infotech	32,190	7.6%	3,800	Pune	

	Rank	Name	Industry	Revenue(in ₹ Crore)	Revenue growth	Profits(in ₹ Crole)	Headquarters	State Controlled
46	47	UltraTech Cement	Cement	32,106	23.7%	2,222	Mumbai	

In [26]:

```
1  df.tail(15)
```

Out[26]:

	Rank	Name	Industry	Revenue(in ₹ Crore)	Revenue growth	Profits(in ₹ Crore)	Headquarters	State Controlled
35	36	ITC Limited	Consumer goods	44,674	1.6%	11,271	Kolkata	
36	37	Bank of India	Banking	44,158	4.8%	–5,961	Mumbai	Yes
37	38	General Insurance Corporation of India	Insurance	43,555	38.9%	3,146	Mumbai	Yes
38	39	Kotak Mahindra Bank	Banking	38,813	14.2%	6,201	Mumbai	
39	40	Union Bank of India	Banking	38,413	0.4%	–5,212	Mumbai	Yes
40	41	Hindustan Unilever	Consumer goods	36,088	7.1%	5,214	Mumbai	
41	42	Vodafone India	Telecommunication	34,855	18.9%	7,766	Mumbai	
42	43	Tata Power	Utilities	33,213	12.8%	2,477	Mumbai	
43	44	Chennai Petroleum Corporation	Oil and gas	33,187	20.0%	927	Chennai	Yes
44	45	Hero MotoCorp	Automotive	33,141	12.9%	3,720	New Delhi	
45	46	Tech Mahindra	Infotech	32,190	7.6%	3,800	Pune	
46	47	UltraTech Cement	Cement	32,106	23.7%	2,222	Mumbai	
47	48	Petronet LNG	Oil and gas	30,949	23.9%	2,110	New Delhi	Yes
48	49	Power Grid Corporation of India	Utilities	30,908	16.1%	8,198	Gurgaon	Yes
49	50	Bajaj Finserv	Financials	30,599	24.9%	2,741	Pune	

In [28]:

```
1  df.to_csv('indian_companies.csv')
```

In [ ]:

```
1
```