

**web scrapping** if api is not provided

we are gona scrap data from AmbitionBox website <https://www.ambitionbox.com/list-of-companies?page=1> we will gona hit with page number

company name ,

```
import pandas as pd
import requests
from bs4 import BeautifulSoup
```

```
requests.get('https://www.ambitionbox.com/list-of-companies?page=1').text
```

```
'<HTML><HEAD>\n<TITLE>Access Denied</TITLE>\n</HEAD><BODY>\n<H1>Access Denied</H1>\n\nYou don\'t have permission to access "http&#58;&#47;&#47;www&#46;ambitionbox&#46;com&#47;list&#45;of&#45;companies&#63;" on this server.<P>\nReference&#32;&#35;18&#46:54403617&#46:1668137208&#46:73ce1f5h\n</BODY>\n</HTML>\n'
```

#file is being protected by robot.txt from crawling data from https request :

#sol disguise header as a humanbeing

#if response code is 403

```
#headers={'User-Agent':'Mozilla/5.0 (Windows NT 6.3; Win 64 ; x64) Apple WeKit /537.36(KHTM
#-requests.get('url',headers=headers).text
```

```
headers={'User-Agent':'Mozilla/5.0 (Windows NT 6.3; Win 64 ; x64) Apple WeKit /537.36(KHTM
webpage=requests.get('https://www.ambitionbox.com/list-of-companies?page=1',headers=header
```

```
soup=BeautifulSoup(webpage,'lxml')
```

#for understandable html page

```
print(soup.prettify())
```

```
<!DOCTYPE html>
<html data-n-head="%7B%22lang%22:%7B%22ssr%22:%22en%22%7D%7D" data-n-head-ssr="" 1
  <head>
    <meta charset="utf-8"/>
    <meta content="width=device-width,initial-scale=1,minimum-scale=1" name="viewport"/>
    <meta content="IE=edge" http-equiv="X-UA-Compatible"/>
    <title>
      List of companies - 747.9k companies | AmbitionBox
    </title>
    <meta content="2022 AmbitionBox" data-n-head="ssr" name="copyright"/>
    <meta content="1 day" data-n-head="ssr" name="revisit-after"/>
    <meta content="AmbitionBox" data-n-head="ssr" name="application-name"/>
    <meta content="EN" data-n-head="ssr" name="content-language"/>
    <meta content="462822053404-hphug4pkahqljh2tc96g35at47o4isv2.apps.googleusercontent.com" data-n-head="ssr" property="fb:app_id"/>
    <meta content="712617688793459" data-n-head="ssr" property="fb:app_id"/>
    <meta content="#ffffff" data-n-head="ssr" name="theme-color"/>
    <meta content="#ffffff" data-n-head="ssr" name="msapplication-navbutton-color"/>
    <meta content="default" data-n-head="ssr" name="apple-mobile-web-app-status-bar-style"/>
    <meta content="100001438127755,100000444923785" data-n-head="ssr" property="fb:app_id"/>
    <meta content="AmbitionBox" data-n-head="ssr" property="og:site_name"/>
```

```
<meta content="website" data-n-head="ssr" property="og:type"/>
<meta content="summary_large_image" data-n-head="ssr" name="twitter:card"/>
<meta content="ambitionbox.com" data-n-head="ssr" name="twitter:domain"/>
<meta content="www.ambitionbox.com" data-n-head="ssr" name="author"/>
<meta content="document" data-n-head="ssr" name="resource-type"/>
<meta content="GLOBAL" data-n-head="ssr" name="distribution"/>
<meta content="general" data-n-head="ssr" name="rating"/>
<meta content="no-cache" data-n-head="ssr" name="pragma"/>
<meta content="Company:Reviews,Salaries,Benefits,Overview,Interview Questions" data-n-head="ssr" name="description"/>
<meta content="https://static.ambitionbox.com/alpha/img/supporting-images/browse" data-n-head="ssr" name="image"/>
<meta content="yes" data-n-head="ssr" name="mobile-web-app-capable"/>
<meta content="AmbitionBox" data-n-head="ssr" name="apple-mobile-web-app-title"/>
<meta content="ANupXnJc0OGVOEFKYteoaBHgOPipYkAdjm3iRKvcdgI" data-n-head="ssr" name="apple-mobile-web-app-status-bar-style"/>
<meta content="List of companies - 747.9k companies | AmbitionBox" data-hid="og:title" data-n-head="ssr" data-cs="2" data-kind="parent" data-bbox="111 356 989 378" data-kind="ghost">
```

```

class="company-name bold-title-1" title="HDFC Bank">
    HDFC Bank
</h2>, <h2
class="company-name bold-title-1" title="Wipro">
    Wipro
</h2>, <h2
class="company-name bold-title-1" title="Infosys">
    Infosys
</h2>, <h2
class="company-name bold-title-1" title="Capgemini">
    Capgemini
</h2>, <h2
class="company-name bold-title-1" title="Tech Mahindra">
    Tech
Mahindra
</h2>, <h2
class="company-name bold-title-1" title="Genpact">
    Genpact
</h2>, <h2
class="company-name bold-title-1" title="HCLTech">
    HCLTech
</h2>, <h2
class="company-name bold-title-1" title="Amazon">
    Amazon
</h2>, <h2
class="company-name bold-title-1" title="Axis Bank">
    Axis Bank
</h2>, <h2
class="company-name bold-title-1" title="IBM">
    IBM
</h2>, <h2
class="company-name bold-title-1" title="Concentrix Corporation">
    Concentrix
Corpo...
</h2>, <h2
class="company-name bold-title-1" title="Reliance jio">
    Reliance
jio
</h2>, <h2
class="company-name bold-title-1" title="Larsen & Toubro Limited">
    Larsen
& Toubro ...
</h2>, <h2
class="company-name bold-title-1" title="Vodafone Idea">
    Vodafone
Idea

```

```

for i in soup.find_all('h2'):
    print(i.text.strip())                #strip unwanted text

```

```

TCS
Accenture
Cognizant
ICICI Bank
HDFC Bank
Wipro
Infosys
Capgemini

```

Tech Mahindra  
 Genpact  
 HCLTech  
 Amazon  
 Axis Bank  
 IBM  
 Concentrix Corpo...  
 Reliance jio  
 Larsen & Toubro ...  
 Vodafone Idea  
 HDB Financial Se...  
 Teleperformance  
 Kotak Mahindra B...  
 Reliance Industr...  
 Bharti Airtel  
 Reliance Retail  
 Deloitte  
 Tata Motors  
 BYJU'S  
 WNS  
 Mahindra & Mahin...  
 Flipkart

#ratings

#rating and almost all stores in p class in html page ( check using inspect on webpage)

```
soup.find_all('p')
```

```
[<p class="text" style="color:#040a1e;">
    AmbitionBox
    </p>,
    <p class="subNavDescription">Discover best places to work</p>,
    <p class="subNavDescription">Compare & find best workplace</p>,
    <p class="subNavDescription">Read reviews for 6L+ companies</p>,
    <p class="subNavDescription">Rate your former or current company</p>,
    <p class="subNavDescription">Discover salaries for 6L+ companies</p>,
    <p class="subNavDescription">Calculate your take home salary</p>,
    <p class="subNavDescription">Check your market value</p>,
    <p class="subNavDescription">Help other jobseekers</p>,
    <p class="subNavDescription">Read interviews for 40K+ companies</p>,
    <p class="subNavDescription">Interviews questions for 1K+ colleges</p>,
    <p class="subNavDescription">Contribute your interview questions</p>,
    <p class="typeahead-title bold-display-1" style="line-
height:normal;color:#1e223c;">
        Discover Best Places to Work!
    </p>,
    <p class="subtitle" style="color:#1e223c;">Company reviews. Salaries. Interviews.
Jobs.</p>,
    <p data-v-6a0700ed="">About Company</p>,
    <p class="sbold-section-header subtitle">
        7,47,874 unique
        companies found
    </p>,
    <p class="sbold-list-header">Sort By: </p>,
    <p class="bold-list-header show-flex">
        Popular
        <i class="icon-chevron-
right"></i></p>,
    <p class="rating badge-large rating-35"><i class="icon icon-star"></i>
```

3.9

```

</p>,
<p data-v-57e129ff="">#6 Best Mega Company - 2021</p>,
<p class="infoEntity sbold-list-header"><i class="icon-domain"></i>
    Public
    </p>,
<p class="infoEntity sbold-list-header"><i class="icon-pin-drop"></i>
    Mumbai,Maharashtra + 276
more
    </p>,
<p class="infoEntity sbold-list-header"><i class="icon-access-time"></i>
    54 years old
    </p>,
<p class="infoEntity sbold-list-header"><i class="icon-supervisor-account"></i>
    1 Lakh+ Employees (India)
    </p>,
<p class="description body-small" itemprop="description">
    We ensure the highest levels of certainty
and satisfaction through a deep-set commitment to our clients, comprehensive
industry expertise and a global network of innovation and delivery centers.

    We function as a full stakeholder to business, offering a consulting-led approach
with an integrated portfolio of technology led solutions that encompass the entire
Enterprise value chain. Our Customer-centric Engagement Model defines how we do
engage with you, offering specialized services and solutions that meet the
distinct needs of your business.

```

```

for i in soup.find_all('p'):    #p tags are used for all except company name
    print(i.text)

```

### AmbitionBox

Discover best places to work  
 Compare & find best workplace  
 Read reviews for 6L+ companies  
 Rate your former or current company  
 Discover salaries for 6L+ companies  
 Calculate your take home salary  
 Check your market value  
 Help other jobseekers  
 Read interviews for 40K+ companies  
 Interviews questions for 1K+ colleges  
 Contribute your interview questions

Discover Best Places to Work!

Company reviews. Salaries. Interviews. Jobs.  
 About Company

7,47,874 unique  
 companies found

Sort By:

Popular

## #6 Best Mega Company - 2021

Public

Mumbai,Maharashtra + 276 m

54 years old

1 Lakh+ Employees (India)

We ensure the highest levels of certainty

We function as a full stakeholder to business, offering a consulting-led approach

We build bespoke teams around your domain and technology requirements drawn from o

Our organization structure is domain led and empowered to help provide Customers a

We have made significant investments in Digital platforms and products spanning Te

4.1

```
soup.find_all('p',class_='rating')
```

```
[<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.9
</p>,
<p class="rating badge-large rating-4"><i class="icon icon-star"></i>
4.1
</p>,
<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.9
</p>,
<p class="rating badge-large rating-4"><i class="icon icon-star"></i>
4.0
</p>,
<p class="rating badge-large rating-4"><i class="icon icon-star"></i>
4.0
</p>,
<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.9
</p>,
<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.9
</p>,
<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.9
</p>,
<p class="rating badge-large rating-35"><i class="icon icon-star"></i>
3.7
</p>,
<p class="rating badge-large rating-4"><i class="icon icon-star"></i>
```

```

                                </p>, 4.0
    <p class="rating badge-large rating-35"><i class="icon icon-star"></i>
                                </p>, 3.8
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.3
    <p class="rating badge-large rating-35"><i class="icon icon-star"></i>
                                </p>, 3.9
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.2
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.1
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.0
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.1
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.2
    <p class="rating badge-large rating-4"><i class="icon icon-star"></i>
                                </p>, 4.0
    <p class="rating badge-large rating-35"><i class="icon icon-star"></i>

```

```

for i in soup.find_all('p',class_='rating'):
    print (i.text.strip())

```

```

3.9
4.1
3.9
4.0
4.0
3.9
3.9
3.9
3.7
4.0
3.8
4.3
3.9
4.2
4.1
4.0
4.1
4.2
4.0
3.6
3.9
4.1
4.1
4.1
4.1

```

4.1  
3.5  
3.7  
3.9  
4.2

#reviews

soup.find\_all('a',class\_='review-count')

```
[<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/tcs-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/accenture-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/cognizant-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/icici-bank-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/hdfc-bank-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/wipro-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/infosys-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/capgemini-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/tech-mahindra-reviews">
Reviews)
</a>,
<a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/genpact-reviews">
Reviews)
</a>]
```

(49.9k  
(32.5k  
(29.7k  
(35k  
(40k  
(28.8k  
(28.3k  
(24.1k  
(21.9k  
(21.1k



```

    <a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/hcl-technologies-reviews">
Reviews)
</a>,
    <a class="review-count sbold-Labels"
href="https://www.ambitionbox.com/reviews/amazon-reviews">
(18.3k

```

```

for i in soup.find_all('a',class_='review-count'):
    print(i.text.strip())

```

```

(49.9k Reviews)
(32.5k Reviews)
(29.7k Reviews)
(35k Reviews)
(40k Reviews)
(28.8k Reviews)
(28.3k Reviews)
(24.1k Reviews)
(21.9k Reviews)
(21.1k Reviews)
(18.6k Reviews)
(18.3k Reviews)
(19.1k Reviews)
(15.7k Reviews)
(15.2k Reviews)
(15.1k Reviews)
(24.9k Reviews)
(13.9k Reviews)
(13.5k Reviews)
(14.5k Reviews)
(13.8k Reviews)
(44k Reviews)
(12.9k Reviews)
(15.7k Reviews)
(10.3k Reviews)
(12.5k Reviews)
(12k Reviews)
(7k Reviews)
(39.8k Reviews)
(11.6k Reviews)

```

```

#type of company
soup.find_all('p',class_='rating')

```

### best to wrap data company wise

```

company=soup.find_all('div',class_='company-content-wrapper')

```

```

len(company)

```

30

```
name=[]  
for i in company:  
  
    print(i.find('h2').text.strip())
```

```
TCS  
Accenture  
Cognizant  
ICICI Bank  
HDFC Bank  
Wipro  
Infosys  
Capgemini  
Tech Mahindra  
Genpact  
HCLTech  
Amazon  
Axis Bank  
IBM  
Concentrix Corpo...  
Reliance jio  
Larsen & Toubro ...  
Vodafone Idea  
HDB Financial Se...  
Teleperformance  
Kotak Mahindra B...  
Reliance Industr...  
Bharti Airtel  
Reliance Retail  
Deloitte  
Tata Motors  
BYJU'S  
WNS  
Mahindra & Mahin...  
Flipkart
```

```
#company name  
name=[]  
for i in company:  
  
    name.append(i.find('h2').text.strip())
```

```
name
```

```
['TCS',  
 'Accenture',  
 'Cognizant',  
 'ICICI Bank',  
 'HDFC Bank',  
 'Wipro',  
 'Infosys',  
 'Capgemini',  
 'Tech Mahindra',  
 'Genpact',  
 'HCLTech',  
 'Amazon',  
 'Axis Bank',
```

```
'IBM',
'Concentrix Corpo...',
'Reliance jio',
'Larsen & Toubro ...',
'Vodafone Idea',
'HDB Financial Se...',
'Teleperformance',
'Kotak Mahindra B...',
'Reliance Industr...',
'Bharti Airtel',
'Reliance Retail',
'Deloitte',
'Tata Motors',
"BYJU'S",
'WNS',
'Mahindra & Mahin...',
'Flipkart']
```

```
#ratings
```

```
#company name
```

```
name=[]
```

```
rating=[]
```

```
for i in company:
```

```
    name.append(i.find('h2').text.strip())
```

```
    rating.append(i.find('p',class_='rating').text.strip())#p have multiple items in a singl
```

```
rating
```

```
['3.9',
'4.1',
'3.9',
'4.0',
'4.0',
'3.9',
'3.9',
'3.9',
'3.7',
'4.0',
'3.8',
'4.3',
'3.9',
'4.2',
'4.1',
'4.0',
'4.1',
'4.2',
'4.0',
'3.6',
'3.9',
'4.1',
'4.1',
'4.1',
'4.1',
'4.1',
'3.5',
'3.7',
```

```
'3.9',
'4.2']
```

```
#review
#ratings
#company name
name=[]
rating=[]
reviews=[]
for i in company:

    name.append(i.find('h2').text.strip())
    rating.append(i.find('p',class_='rating').text.strip())#p have multiple items in a single
    reviews.append(i.find('a',class_='review-count').text.strip())
```

```
reviews
```

```
['(49.9k Reviews)',
'(32.5k Reviews)',
'(29.7k Reviews)',
'(35k Reviews)',
'(40k Reviews)',
'(28.8k Reviews)',
'(28.3k Reviews)',
'(24.1k Reviews)',
'(21.9k Reviews)',
'(21.1k Reviews)',
'(18.6k Reviews)',
'(18.3k Reviews)',
'(19.1k Reviews)',
'(15.7k Reviews)',
'(15.2k Reviews)',
'(15.1k Reviews)',
'(24.9k Reviews)',
'(13.9k Reviews)',
'(13.5k Reviews)',
'(14.5k Reviews)',
'(13.8k Reviews)',
'(44k Reviews)',
'(12.9k Reviews)',
'(15.7k Reviews)',
'(10.3k Reviews)',
'(12.5k Reviews)',
'(12k Reviews)',
'(7k Reviews)',
'(39.8k Reviews)',
'(11.6k Reviews)']
```

```
#def of a company
company[0]
```

```
<div class="company-content-wrapper"><div class="company-content"><div
class="company-logo"></div> <div class="company-info-wrapper"><div
class="company-info"><div class="left"><a href="/overview/tcs-overview"><h2
class="company-name bold-title-1" title="TCS">
TCS
</h2></a> <div
class="rating-wrapper"><p class="rating badge-large rating-35"><i class="icon
icon-star"></i>
3.9
</p> <a
class="review-count sbold-labels" href="https://www.ambitionbox.com/reviews/tcs-
reviews">
(49.9k
Reviews)
</a></div></div>
<button class="ab_btn follow-btn invert round"><span class="ctas-btn-
medium">Follow</span></button></div> <div class="leaf-list-wrapper noscrollbars-sm
mg-btm-16 mg-top-4" data-v-1018e3e7=""><div class="award-leaf" data-v-1018e3e7=""
data-v-57e129ff="">  <p data-v-57e129ff="">#6 Best Mega Company
- 2021</p> </div><!-- --> <!-- --></div> <div class="company-basic-
info"><p class="infoEntity sbold-list-header"><i class="icon-domain"></i>
Public
</p> <p class="infoEntity sbold-
list-header"><i class="icon-pin-drop"></i>
Mumbai,Maharashtra + 276
more
</p> <p class="infoEntity sbold-
list-header"><i class="icon-access-time"></i>
54 years old
</p> <p class="infoEntity sbold-
list-header"><i class="icon-supervisor-account"></i>
1 Lakh+ Employees (India)
</p></div> <ul class="chips-
block"><li><a class="ab_chip body-medium" data-filter-name="chips_Company-
Tags_bpo" href="/bpo-companies-in-india" title="BPO companies in India">
BPO
</a></li><li><a
class="ab_chip body-medium" data-filter-name="chips_Company-Tags_it-services-and-
consulting" href="/it-services-and-consulting-companies-in-india" title="IT
Services & Consulting companies in India">
IT Services & Consulting
</a></li><li><a
class="ab_chip body-medium" data-filter-name="chips_Company-Tags_forbesglobal2000"
href="/forbesglobal2000-companies-in-india" title="Forbes Global 2000 companies in
India">
Forbes Global 2000

```

```
company[0].find_all('p',class_='infoEntity')[0].text.strip()
```

```
'Public '
```

```
company[0].find_all('p',class_='infoEntity')[1].text.strip()
```

```
'Mumbai,Maharashtra + 276 more'
```

```
company[0].find_all('p',class_='infoEntity')[2].text.strip()
```

```
'54 years old'
```

```
company[0].find_all('p',class_='infoEntity')[3].text.strip()
```

```
'1 Lakh+ Employees (India)'
```

```
#review
```

```
#ratings
```

```
#company name
```

```
name=[]
```

```
rating=[]
```

```
reviews=[]
```

```
ctype=[]
```

```
hq=[]
```

```
old=[]
```

```
employees=[]
```

```
desc=[]
```

```
for i in company:
```

```
    name.append(i.find('h2').text.strip())
```

```
    rating.append(i.find('p',class_='rating').text.strip())#p have multiple items in a singl
```

```
    reviews.append(i.find('a',class_='review-count').text.strip())
```

```
    ctype.append(i.find_all('p',class_='infoEntity')[0].text.strip())
```

```
    hq.append(i.find_all('p',class_='infoEntity')[1].text.strip())
```

```
    old.append(i.find_all('p',class_='infoEntity')[2].text.strip())
```

```
    employees.append(i.find_all('p',class_='infoEntity')[3].text.strip())
```

```
    desc.append(i.find_all('p',class_='description')[0].text.strip())
```

```
desc[0]
```

```
'We ensure the highest levels of certainty and satisfaction through a deep-set commi
tment to our clients, comprehensive industry expertise and a global network of innov
ation and delivery centers.\n\nWe function as a full stakeholder to business, offeri
ng a consulting-led approach with an integrated portfolio of technology led solution
s that encompass the entire Enterprise value chain. Our Customer-centric Engagement
Model defines how we do engage with you, offering specialized services and solutions
that meet the distinct needs of your business.\n\nWe build bespoke teams around your
domain and technology requirements drawn from our talent pool of over 488,649 global
```

```
#review
```

```
#ratings
```

```
#company name
```

```
name=[]
```

```
rating=[]
```

```
reviews=[]
```

```
ctype=[]
```

```
hq=[]
```

```

old=[]
employees=[]
desc=[]
for i in company:

    name.append(i.find('h2').text.strip())
    rating.append(i.find('p',class_='rating').text.strip())#p have multiple items in a singl
    reviews.append(i.find('a',class_='review-count').text.strip())
    ctype.append(i.find_all('p',class_='infoEntity')[0].text.strip())
    hq.append(i.find_all('p',class_='infoEntity')[1].text.strip())
    old.append(i.find_all('p',class_='infoEntity')[2].text.strip())
    employees.append(i.find_all('p',class_='infoEntity')[3].text.strip())
    desc.append(i.find_all('p',class_='description')[0].text.strip())

d={'name':name,'rating':rating,'reviews':reviews,'type':ctype,'HeadQuater':hq,'Old':old,'E
df=pd.DataFrame(d)

```

```
df.head()
```

	name	rating	reviews	type	HeadQuater	Old	Employees	Descripti
0	TCS	3.9	(49.9k Reviews)	Public	Mumbai,Maharashtra + 276 more	54 years old	1 Lakh+ Employees (India)	We ensu the highe levels certainty ai
1	Accenture	4.1	(32.5k Reviews)	Public	Dublin + 141 more	33 years old	1 Lakh+ Employees (India)	Accenture a glob professor services cc
2	Cognizant	3.9	(29.7k Reviews)	Private	Teaneck. New Jersev. + 98 more	28 years	1 Lakh+ Employees	At Cogniza we gi organizatio

```

# apply it in all pages
import numpy as np
final=pd.DataFrame()
for j in range(1,333):
    headers={'User-Agent':'Mozilla/5.0 (Windows NT 6.3; Win 64 ; x64) Apple WeKit /537.36(KH
    webpage=requests.get('https://www.ambitionbox.com/list-of-companies?page={}'.format(j),h
    soup=BeautifulSoup(webpage,'lxml')
    company=soup.find_all('div',class_='company-content-wrapper')
    name=[]
    rating=[]
    reviews=[]
    ctype=[]
    hq=[]
    how_old=[]
    no_of_employee=[]
    desc=[]

    for i in company:

```

```
try:
    name.append(i.find('h2').text.strip())
except:
    name.append(np.nan)

try:
    rating.append(i.find('p',class_='rating').text.strip())
except:
    rating.append(np.nan)

try:
    reviews.append(i.find('a' , class_='review-count').text.strip())
except:
    reviews.append(np.nan)

try:
    ctype.append(i.find_all('p',class_='infoEntity')[0].text.strip())
except:
    ctype.append(np.nan)
try:
    hq.append(i.find_all('p',class_='infoEntity')[1].text.strip())
except:
    hq.append(np.nan)

try:
    how_old.append(i.find_all('p',class_='infoEntity')[2].text.strip())
except:
    how_old.append(np.nan)
try:
    no_of_employee.append(i.find_all('p',class_='infoEntity')[3].text.strip())
except:
    no_of_employee.append(np.nan)

try:
    desc.append(i.find_all('p',class_='description')[0].text.strip())
except:
    desc.append(np.nan)

df=pd.DataFrame({'name':name,
                 'rating':rating,
                 'reviews':reviews,
                 'company_type':ctype,
                 'Head_Quarters':hq,
                 'Company_Age':how_old,
                 'No_of_Employee':no_of_employee,
                 'Description':desc
                 })
final=final.append(df,ignore_index=True)

df.head()
```



	name	rating	reviews	company_type	Head_Quarters	Company_Age	No_of_Employ
0	QNFS	4.6	(53 Reviews)	Doha + 3 more	50 years old	501-1k Employees (Global)	N
1	R.K. Group	3.5	(53 Reviews)	Private	Kolkata, West Bengal + 20 more	38 years old	51-2 Employees (Inc

1 | 1

df.shape

(30, 8)

final.shape

(9958, 8)

final.head()

	name	rating	reviews	company_type	Head_Quarters	Company_Age	No_of_Ei
0	TCS	3.9	(49.9k Reviews)	Public	Mumbai, Maharashtra + 276 more	54 years old	Err
1	Accenture	4.1	(32.5k Reviews)	Public	Dublin + 141 more	33 years old	Err
2	Cognizant	3.9	(29.7k Reviews)	Private	Teaneck, New Jersey + 98 more	28 years old	Err

final.to\_csv('India\_top\_~9000\_companies')

Double-click (or enter) to edit

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 3:02 PM

