## web scrapping if api is not provided

we are gona scrap data from AmbitionBox website <a href="https://www.ambitionbox.com/list-of-companies?page=1">https://www.ambitionbox.com/list-of-companies?page=1</a> 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://www.ambitionbox.c om/list-of-companies?" on this server.<P>\nReference&#32;&#35;18&#4 6:54403617%#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(KHT #-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="viewpor"</pre> <meta content="IE=edge" http-equiv="X-UA-Compatible"/> <title> List of companies - 747.9k companies | AmbitionBox <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.googleusercont</pre> <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<meta content="100001438127755,100000444923785" data-n-head="ssr" property="fb:ac</pre>

<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" d</pre>
       <meta content="https://static.ambitionbox.com/alpha/img/supporting-images/browse</pre>
       <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"/</pre>
       <meta content="ANupXnJc00GV0EFKYteoaBHg0PipYkAdjm3iRKvcdgI" data-n-head="ssr" nar</pre>
       <meta content="List of companies - 747.9k companies | AmbitionBox" data-hid="og:</pre>
       <meta content="List of companies - 747.9k companies | AmbitionBox" data-hid="twi-</pre>
       <meta content="View list of companies in India with in-depth information on revi</pre>
       <meta content="View list of companies in India with in-depth information on revi</pre>
       <meta content="View list of companies in India with in-depth information on revious."</pre>
       <meta content="https://static.ambitionbox.com/static/list-of-companies.jpg" data</pre>
       <meta content="https://static.ambitionbox.com/static/list-of-companies.jpg" data</pre>
       <meta content="ALL" data-n-head="ssr" name="robots"/>
       <meta content="https://www.ambitionbox.com/list-of-companies" data-n-head="ssr"</pre>
       <meta content="https://www.ambitionbox.com/list-of-companies" data-n-head="ssr"</pre>
       <link data-n-head="ssr" href="https://static.ambitionbox.com/alpha/img/supporting</pre>
       <link crossorigin="anonymous" data-n-head="ssr" href="https://static.ambitionbox</pre>
       <link crossorigin="anonymous" data-n-head="ssr" href="https://static.ambitionbox</pre>
       <link crossorigin="anonymous" data-n-head="ssr" href="https://cdnjs.cloudflare.c</pre>
       <link data-n-head="ssr" href="https://cdnjs.cloudflare.com/ajax/libs/lightslider</pre>
       <link data-n-head="ssr" href="https://www.ambitionbox.com/list-of-companies" rel</pre>
       <link data-n-head="ssr" href="https://www.ambitionbox.com/list-of-companies" rel</pre>
       <link data-n-head="ssr" href="https://www.ambitionbox.com/list-of-companies?page</pre>
# inspect the webpage and check html script format : https://www.ambitionbox.com/list-of-c
soup.find all('h1')
                          #company name stored in h1 headder in html page
     [<h1 class="bold-title-l title">List of companies in India</h1>]
soup.find_all('h2')
                      #h2 tage contain company name
     [<h2 class="company-name bold-title-1" title="TCS">
                                                                                 TCS
                                                                         </h2>, <h2
     class="company-name bold-title-1" title="Accenture">
                                                                                 Accenture
                                                                         </h2>, <h2
     class="company-name bold-title-1" title="Cognizant">
                                                                                 Cognizant
                                                                         </h2>, <h2
     class="company-name bold-title-1" title="ICICI Bank">
                                                                                 ICICI Bank
```

```
</h2>, <h2
     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'):
                                          #strip unwanted text
```

```
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')
   [
                          AmbitionBox
                      ,
   Discover best places to work,
   Compare & amp; 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,
   height:normal;color:#1e223c;">
            Discover Best Places to Work!
   Company reviews. Salaries. Interviews.
   Jobs.,
   About Company,
   7,47,874 unique
                 companies found
            ,
   Sort By: ,
   Popular
                                    <i class="icon-chevron-</pre>
   right"></i>,
   3.9
```

```
,
#6 Best Mega Company - 2021,
<i class="icon-domain"></i></i></or>
                                        Public Public
                                  \langle p \rangle,
<i class="icon-pin-drop"></i></i>
                                        Mumbai, Maharashtra + 276
more
                                  ,
<i class="icon-access-time"></i></i></or>
                                        54 years old
                                  ,
<i class="icon-supervisor-account"></i></i></or>
                                        1 Lakh+ Employees (India)
                                  ,
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

3.9

#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 be poke 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')

[<i class="icon icon-star"></i>

3.9

,
<i class="icon icon-star"></i>

4.1

<i class="icon icon-star"></i>

3.9

<i class="icon icon-star"></i></i>

4.0

<i class="icon icon-star"></i></i>

4.0

,
<i class="icon icon-star"></i>

crass= racing bauge-range racing-ss ><1 crass= 1con 1con-star ></1>

3.9

<i class="icon icon-star"></i>

3.9

<i class="icon icon-star"></i>

\_

3.9

3.7

<i class="icon icon-star"></i></i>

,

<i class="icon icon-star"></i></i>

4.0

```
,
    <i class="icon icon-star"></i></i></or>
                                                  3.8
   <i class="icon icon-star"></i></i>
                                                  4.3
                                             ,
   <i class="icon icon-star"></i></i>
                                                  3.9
                                             ,
   <i class="icon icon-star"></i></i></or>
                                                  4.2
                                             ,
   <i class="icon icon-star"></i></i></or>
                                                  4.1
   <i class="icon icon-star"></i></i></or>
                                                  4.0
   <i class="icon icon-star"></i></i></or>
                                                  4.1
                                             ,
   <i class="icon icon-star"></i></i></or>
                                                  4.2
   <i class="icon icon-star"></i></i>
                                                  4.0
   <i class="icon icon-star"></i></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"</pre>
href="https://www.ambitionbox.com/reviews/tcs-reviews">
                                                                              (49.9k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/accenture-reviews">
                                                                              (32.5k)
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/cognizant-reviews">
                                                                              (29.7k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/icici-bank-reviews">
                                                                              (35k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/hdfc-bank-reviews">
                                                                              (40k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/wipro-reviews">
                                                                              (28.8k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/infosys-reviews">
                                                                              (28.3k)
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/capgemini-reviews">
                                                                              (24.1k)
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/tech-mahindra-reviews">
                                                                              (21.9k
Reviews)
                                                                     </a>,
 <a class="review-count sbold-Labels"</pre>
href="https://www.ambitionbox.com/reviews/genpact-reviews">
                                                                              (21.1k)
Reviews)
```

</a>,

30

```
<a class="review-count sbold-Labels"</pre>
     href="https://www.ambitionbox.com/reviews/hcl-technologies-reviews">
                                                                                 (18.6k
     Reviews)
                                                                         </a>,
      <a class="review-count sbold-Labels"</pre>
     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)
```

```
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 singl
  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"><img alt="Tata Consultancy Services logo" class="lazy" datasrc="https://static.ambitionbox.com/assets/v2/images/rs:fit:200:200:false:false/bG
height="100" onerror="this.onerror=null; this.src='/static/icons/company-</pre>

```
placeholder.svg';" src="https://static.ambitionbox.com/static/icons/company-
    placeholder.svg" width="100"/></div> <div class="company-info-wrapper"><div</pre>
    class="company-info"><div class="left"><a href="/overview/tcs-overview"><h2</pre>
    class="company-name bold-title-1" title="TCS">
                                                                 </h2></a> <div
    class="rating-wrapper"><i class="icon</pre>
    icon-star"></i>
                                                                         3.9
                                                                  <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-</pre>
    medium">Follow</span></button></div> <div class="leaf-list-wrapper noscrollbars-sm</pre>
    mg-btm-16 mg-top-4" data-v-1018e3e7=""><div class="award-leaf" data-v-1018e3e7=""
    data-v-57e129ff=""><img alt="award" data-v-57e129ff="" height="26"
    src="https://static.ambitionbox.com/static/awards/black-theme/golden-leaf-
    left.png" width="15"/> <img alt="AmbitionBox awards" class="ambox-logo" data-v-</pre>
    57e129ff="" height="17" src="https://static.ambitionbox.com/static/awards/black-
    theme/gold-ambox-logo.svg" width="15"/> #6 Best Mega Company
    - 2021 <img alt="award" data-v-57e129ff="" height="26"
    src="https://static.ambitionbox.com/static/awards/black-theme/golden-leaf-
    right.png" width="15"/></div><!-- --> <!-- --></div> <div class="company-basic-</pre>
    info"><i class="icon-domain"></i></i>
                                                          Public
                                                   <p class="infoEntity sbold-
    list-header"><i class="icon-pin-drop"></i>
                                                         Mumbai, Maharashtra + 276
    more

    list-header"><i class="icon-access-time"></i></i></or>
                                                         54 years old

    list-header"><i class="icon-supervisor-account"></i>
                                                          1 Lakh+ Employees (India)
                                                  </div> <ul class="chips-
    block"><a class="ab chip body-medium" data-filter-name="chips Company-
    Tags_bpo" href="/bpo-companies-in-india" title="BPO companies in India">
                                                          </a><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 & amp; Consulting companies in India">
                                                                 IT Services & amp;
    Consulting
                                                         </a><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()
```

```
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 commitment to our clients, comprehensive industry expertise and a global network of innovation and delivery centers.\n\nWe 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.\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=[]
```

```
11/25/22, 3:10 PM
```

```
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	Description
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 a
1	Accenture	4.1	(32.5k Reviews)	Public	Dublin + 141 more	33 years old	1 Lakh+ Employees (India)	Accenture a glot professior services cc
2	Cognizant	3.9	(29.7k Reviews)	Private	Teaneck. New Jersev. + 98 more	28 years	1 Lakh+ Employees	At Cognizar we gi

```
# 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())
      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()
```

No_of_Employ	Company_Age	Head_Quarters	company_type	reviews	rating	name	
N	501-1k Employees (Global)	50 years old	Doha + 3 more	(53 Reviews)	4.6	QNFS	0
51-2 Employe (Inc	38 years old	Kolkata,West Bengal + 20 more	Private	(53 Reviews)	3.5	R.K. Group	1
1 l al							4

df.shape

(30, 8)

final.shape

(9958, 8)

final.head()

	name	rating	reviews	company_type	Head_Quarters	Company_Age	No_of_E
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 Jersev. + 98 more	28 years old	En

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

×