

揭开互联网公司的神秘面纱，数据解读那些slay整个行业的互联网公司

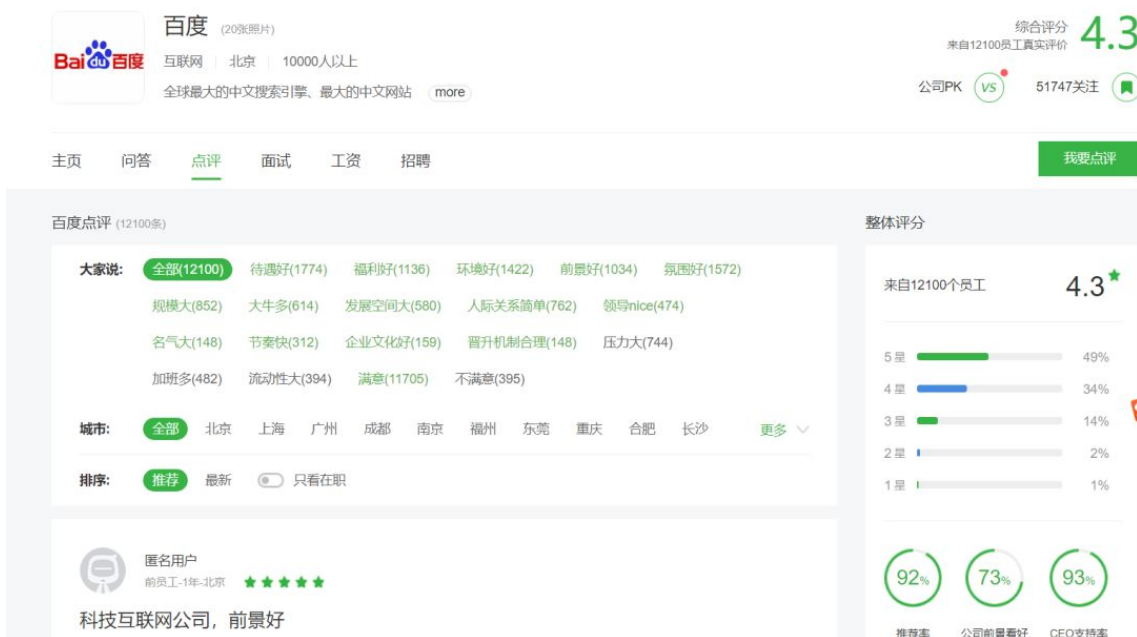
数据森麟

↑
点击这里关注我们



前言：

随着互联网行业的日益兴盛，吸引力越来越多的牛人加入其中，也有许多小伙伴跃跃欲试，想要在互联网的浪潮中大展身手。今天我们通过看准网的数据，帮助大家对各大大互联网公司有一个比较概括的了解。



看准网提供了许多员工对于公司的评价，我们从中提取需要的数据，包括整体评分、面试难度、推荐率、前景看好情况、CEO支持率，代码如下：

```
## 获得信息
```

```
def get_company_info(num, headers):
```

```
## 获得评价数据
```

```
url = 'https://www.kanzhun.com/gsr'+str(num)+' .html?ka=com-blocker1-review'
```

```
js='window.open(""+url+"")'
```

```
driver.execute_script(js)
```

```
time.sleep(5)
```

```
driver.close()
```

```
driver.switch_to_window(driver.window_handles[0])
```

```

bsObj=BeautifulSoup(driver.page_source,"html.parser")

tag=bsObj.find('div',attrs=
{'class':'all_item'}).text.replace('\t','').replace('\n','').replace('(',''
').replace(')',' ').split(' ')

tag=tag[0:len(tag)-1]

this_tag = {tag[i*2]:tag[i*2+1] for i in np.arange(int(len(tag)/2-1))}

this_name = bsObj.find('div',attrs={'class':'co_name t_center'}).text

this_overal = float(bsObj.find('div',attrs={'class':'res_box_star
f_right'}).find('em').text)

points = bsObj.find('ul',attrs={'class':'score_rate
clearfix'}).text.replace('\n',' ').split()

this_recommend = float(points[0][0:2])/100*5

this_future = float(points[2][0:2])/100*5

this_ceo = float(points[4][0:2])/100*5

## 获得CEO头像和公司logo

ceo_pic = bsObj.find('div',attrs=
{'class':'ceo_info'}).find('div').find('img').attrs['src']

ceo_name = bsObj.find('div',attrs={'class':'ceo_info'}).find('p').text

```

```
head_logo = bsObj.find('div', attrs={'class': 'com_logo  
f_left'}).find('img').attrs['src']
```

```
head_loc = 'D:/爬虫/看准/公司logo/'+this_name+'.jpg'
```

```
ceo_loc = 'D:/爬虫/看准/CEOlogo/'+this_name+'.jpg'
```

```
request.urlretrieve(head_logo, head_loc)
```

```
request.urlretrieve(ceo_pic, ceo_loc)
```

```
## 获得面试难度
```

```
url = 'https://www.kanzhun.com/gsm'+str(num)+' .html?ka=com-floater-interview'
```

```
js='window.open(""+url+"")'
```

```
driver.execute_script(js)
```

```
time.sleep(5)
```

```
driver.close()
```

```
driver.switch_to_window(driver.window_handles[0])
```

```
bsObj=BeautifulSoup(driver.page_source, "html.parser")
```

```
req=request.Request(url, headers=headers)
```

```
html=urlopen(req)
```

```
bsObj=BeautifulSoup(html.read(),"html.parser")

    this_difficulty = float(bsObj.find('section',attrs=
{'class':'interview_feel'}).find('em').text)

    this_feeling = bsObj.find('ul',attrs=
{'class':'score_list'}).find_all('span',attrs={'class':'percent'})

    this_feeling = [float(k.text.replace('%','')) for k in this_feeling]

    this_feeling = (this_feeling[0]*5+this_feeling[1]*3+this_feeling[2]*1)/100

    ## 整合数据成为字典

    this_company =
{'name':this_name,'overall':this_overal,'comments':tag[1],'recommend':this_recommend,

'future':this_future,'ceo':this_ceo,'difficulty':this_difficulty,'feeling':this_feeling}

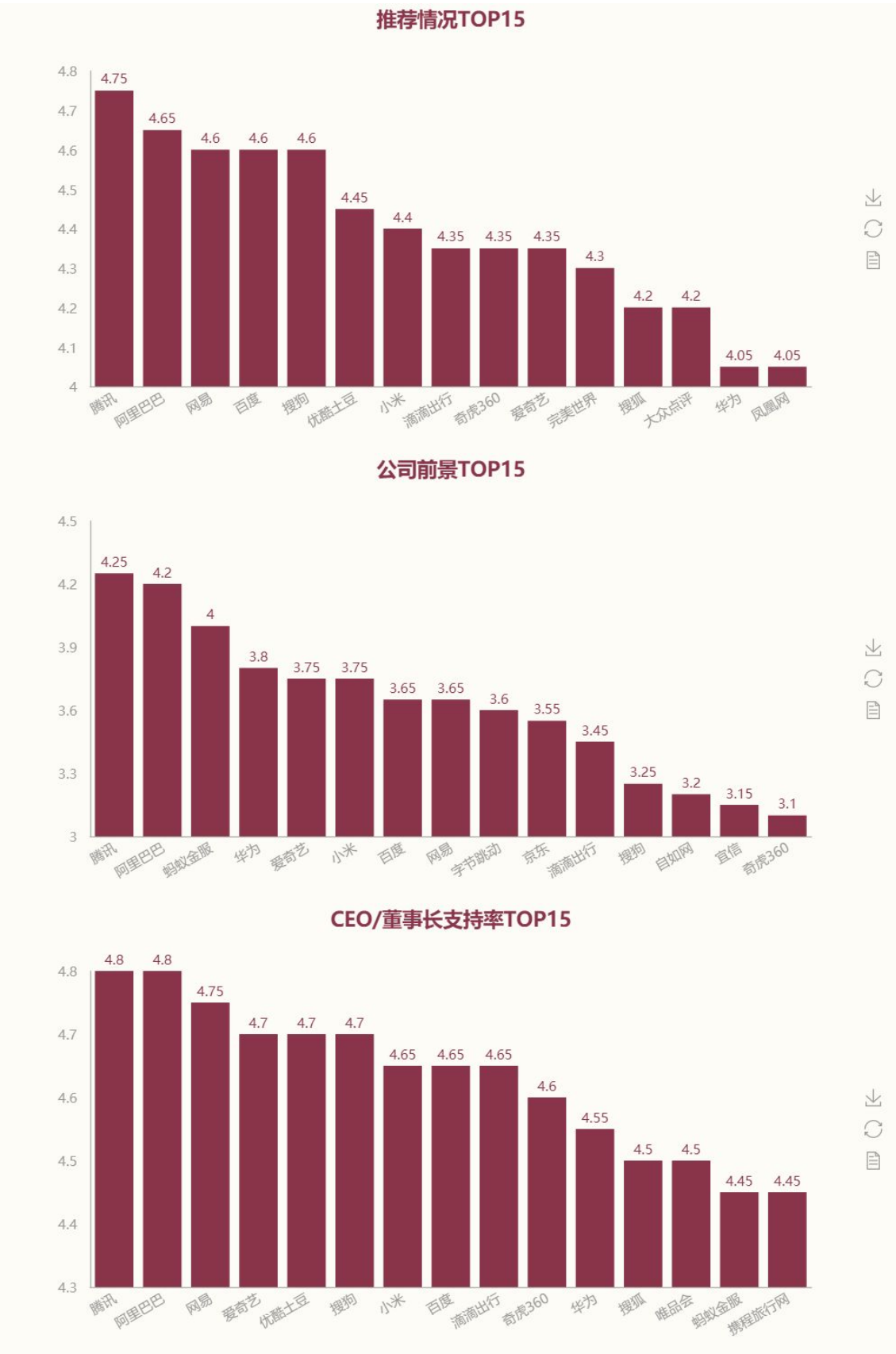
    return this_company,this_tag,this_name
```

我们最终选取了50家互联网公司作为样本进行对比，选取来源主要是结合2018年互联网公司百强榜单和看准网上的实际评价数量，选取的公司logo拼图如下，我们会在第4部分讲解如何将图片进行拼接：

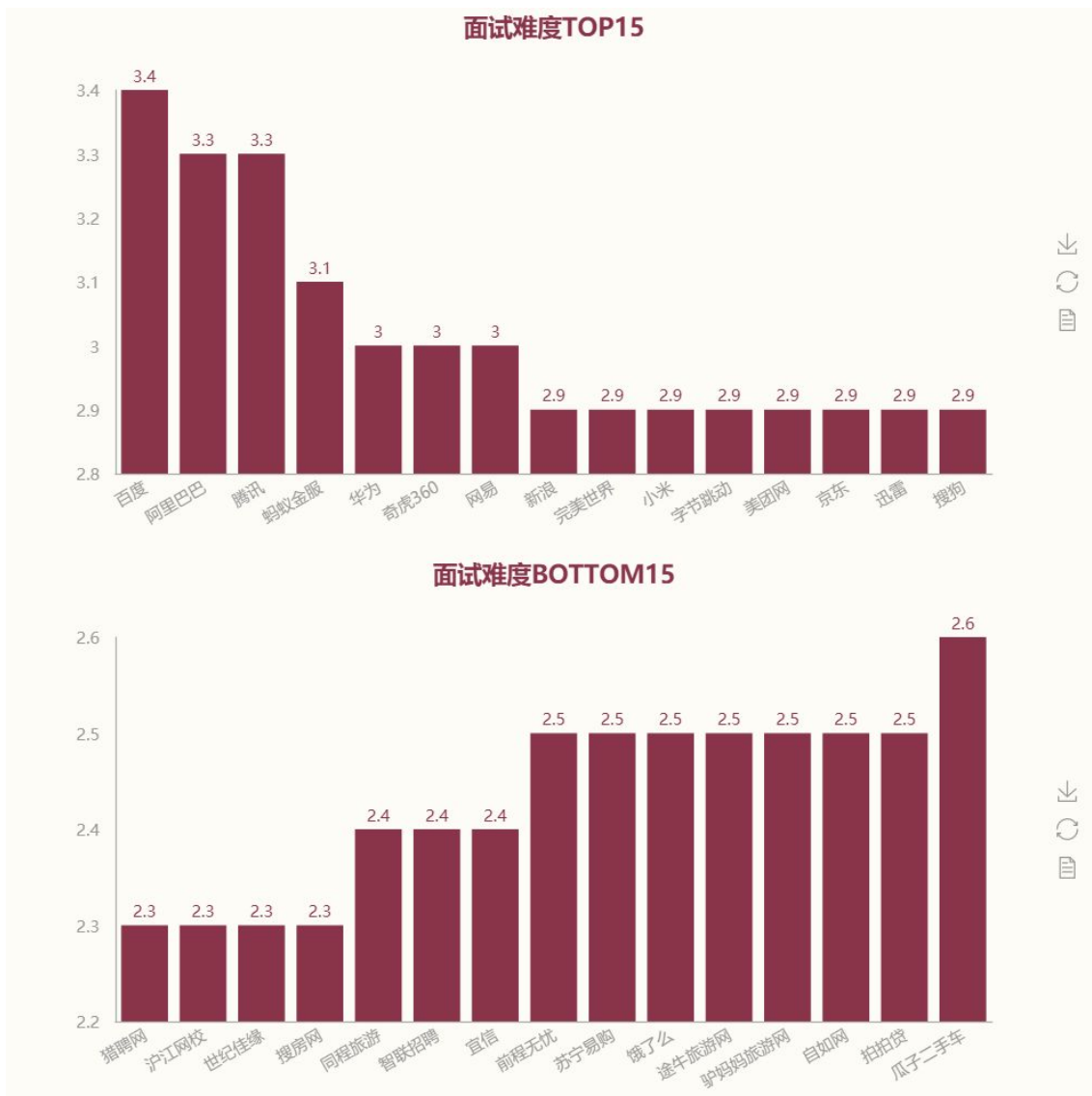


首先对比各项评价指标的TOP15：





可以看到，榜单中BAT在各项排名中都处于十分靠前的位置，网易也占据了多个榜单的靠前位置，腾讯霸占了所有排名的TOP1。下面我们再来看一下面试难度，我们选取了面试难度评分的TOP15和BOTTOM15，该数据仅供参考，根据小编的经验，同一个公司的不同部门不同岗位之间的难度差异也非常大。



在面试难度偏低的一些公司中，有许多非常不错公司，该数据仅仅是一个参考，真正的面试还是要取决于求职者的实际能力，所谓会者不难，难者不会。真正的大牛无论是面对多么困难的面试，依然可以slay全场。

整体评分top15柱形图

```
company=pd.read_excel('company_info.xlsx')
```

```
company_overal = company.sort_values('overall',ascending=False)[0:15]
```

```
attr = company_overal['name']
```

```
v1=round(company_overal['overall'],2)
```



```

bar = Bar("整体评分TOP15", title_pos='center')

bar.use_theme('essos')

bar.add("", attr, v1, is_stack=False, xaxis_rotate=30, yaxis_min=3.7, is_label_show=True,

        xaxis_interval =0, is_splitline_show=False)

bar.render('整体评分TOP15.html')

```

前面我们看的都是各个公司之间的对比，下面我们看一下同一个公司不同维度的情况，我们选取了BAT和TMD作为数据，其他的公司也可以按照同样的方式进行对比，首先看一下BAT：



BAT真的是名副其实的业界标杆，各项指标都slay整个行业，下面我们看一下此前发展势头迅猛的TMD三家公司：



TMD三家公司和行业整体水平相比，也是出于领先地位，可见其还不错的发展势头，最后放上和小编息息相关的三家公司，具体是哪三家，相信了解小编的朋友一定是可以猜出来的：



```
value_avg = [list(company.iloc[:, [1, 3, 4, 5, 6]].mean())]
```

```
value_company0 = [list(company.iloc[0, [1, 3, 4, 5, 6]])]
```

```
value_company1 = [list(company.iloc[1, [1, 3, 4, 5, 6]])]
```

```
value_company2 = [list(company.iloc[2, [1, 3, 4, 5, 6]])]
```

```

c_schema= [{"name": "总体评价", "max": 4.4, "min": 3.2},

            {"name": "推荐度", "max": 4.75, "min": 2.4},

            {"name": "前景看好", "max": 4.25, "min": 1},

            {"name": "CEO/董事长认可度", "max": 4.8, "min": 3},

            {"name": "面试难度", "max": 3.4, "min": 2.3}]

radar = Radar()

radar.use_theme('essos')

radar.config(c_schema=c_schema, shape='circle')

radar.add(company['name'][0], value_company0, item_color="blue",
symbol=None, linewidth=5)

radar.add(company['name'][1], value_company1, item_color="orange",
symbol=None, linewidth=5)

radar.add(company['name'][2], value_company2, item_color="red",
symbol=None, linewidth=5)

radar.add("整体水平", value_avg, item_color="purple", symbol=None, linewidth=5,

            legend_selectedmode='multiple')

radar.render('bat.html')

```

看准网提供了各个公司的logo和各位公司大佬的头像，我们冒昧地利用这些数据进行简单的图片拼接，制作成一副大的合成图。主要原理是利用numpy中的多维数组进行拼接，由于图像本身就可以看做是一个三维数组（彩色）或者一位数组（黑白），所以我们只需利用数组的拼接方法，就可以达到我们的目的。

代码如下：

```
## 拼接公司logo成为5*10的拼图
```

```
i = 0
```

```
for filename in os.listdir("./公司logo"):
```

```
    file_loc = "D:/爬虫/看准/公司logo/"+filename
```

```
    img = mpimg.imread(file_loc)[:,:,:3]
```

```
    img = cv2.resize(img, (180,180), interpolation=cv2.INTER_AREA)
```

```
    if i % 10 == 0:
```

```
        row_img=img
```

```
elif i == 9:

    row_img=np.hstack((row_img, img))

    all_img = row_img

elif i % 10 == 9:

    row_img=np.hstack((row_img, img))

    all_img = np.vstack((all_img, row_img))

else:

    row_img=np.hstack((row_img, img))

    i = i+1

plt.imshow(all_img)

plt.axis('off')

## 拼接大佬头像成为7*7的拼图

i = 0

for filename in os.listdir("./CEOlogo"):

    file_loc = "D:/爬虫/看准/CEOlogo/"+filename
```

```
img = mpimg.imread(file_loc)[:,:,:3]

img = cv2.resize(img, (500,500),interpolation=cv2.INTER_CUBIC)

if i % 7 == 0:

    row_img=img

elif i == 6:

    row_img=np.hstack((row_img,img))

    all_img = row_img

elif i % 7 == 6:

    row_img=np.hstack((row_img,img))

    all_img = np.vstack((all_img,row_img))

else:

    row_img=np.hstack((row_img,img))

i = i+1

plt.imshow(all_img)

plt.axis('off')
```

下面就是我们的效果图，不知道大家是否能一眼就把所有的logo都认全



最后是各位大佬的拼图，不知道大家第一眼看到的是哪位大佬，第一眼看到的大佬，或许就是你未来的老板





欢迎各位读者在留言区与我们互动，聊聊你所了解的互联网公司，或者是你最想去的公司。
如果不方便透露姓名，可以在后台单独留言，我们会以匿名的方式发出来
，欢迎大家一起来分享。

作者介绍：徐麟，目前就职于上海唯品会产品技术中心，哥大统计数据狗，从事数据挖掘&分析工作，喜欢用R&Python玩一些不一样的数据

个人公众号：数据森麟（ID:shujusenlin），知乎同名专栏作者。



END

[爬了下知乎神回复，笑死人了~](#)

[12306:妈妈再也不用担心我抢不到票了\(2\)](#)

[官宣：我最好的东西](#)

[LeetCode005:最长公共前缀](#)

[爬取猫眼信息数据分析可视化](#)

扫码加X先生微信进学习交流群[福利多多]



温馨提示

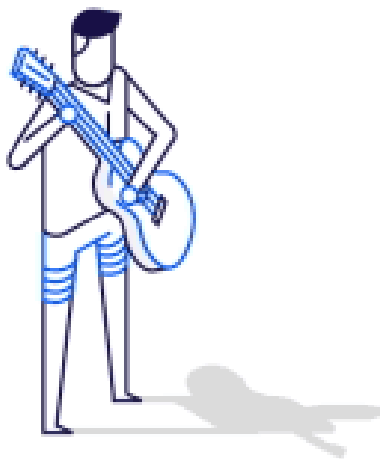
欢迎大家转载，转发，留言，点赞支持X先生。

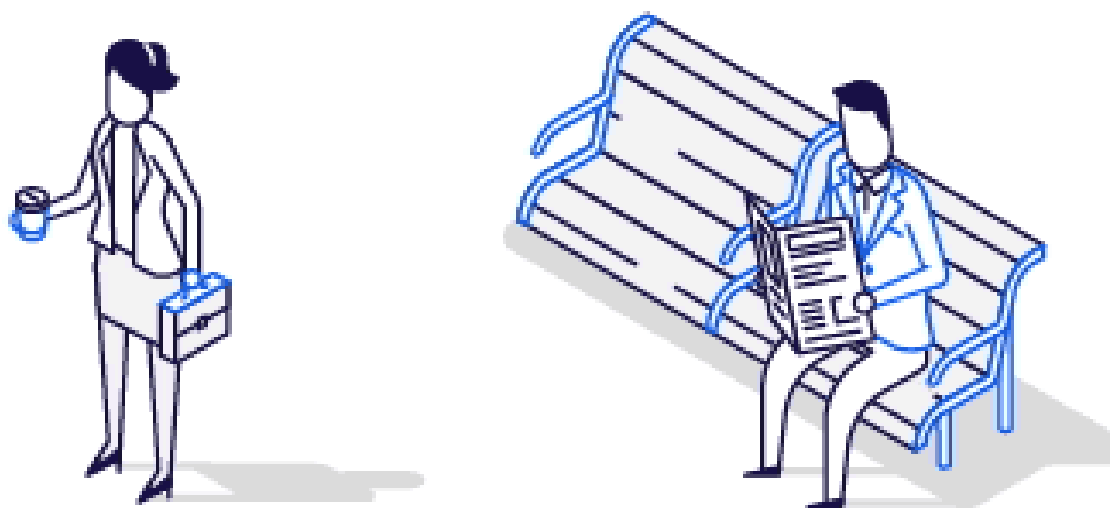
文末广告点一下也是对X先生莫大的支持。

做知识的传播者，随手转发。



更多精彩推荐，请关注我们





文章转载自公众号

