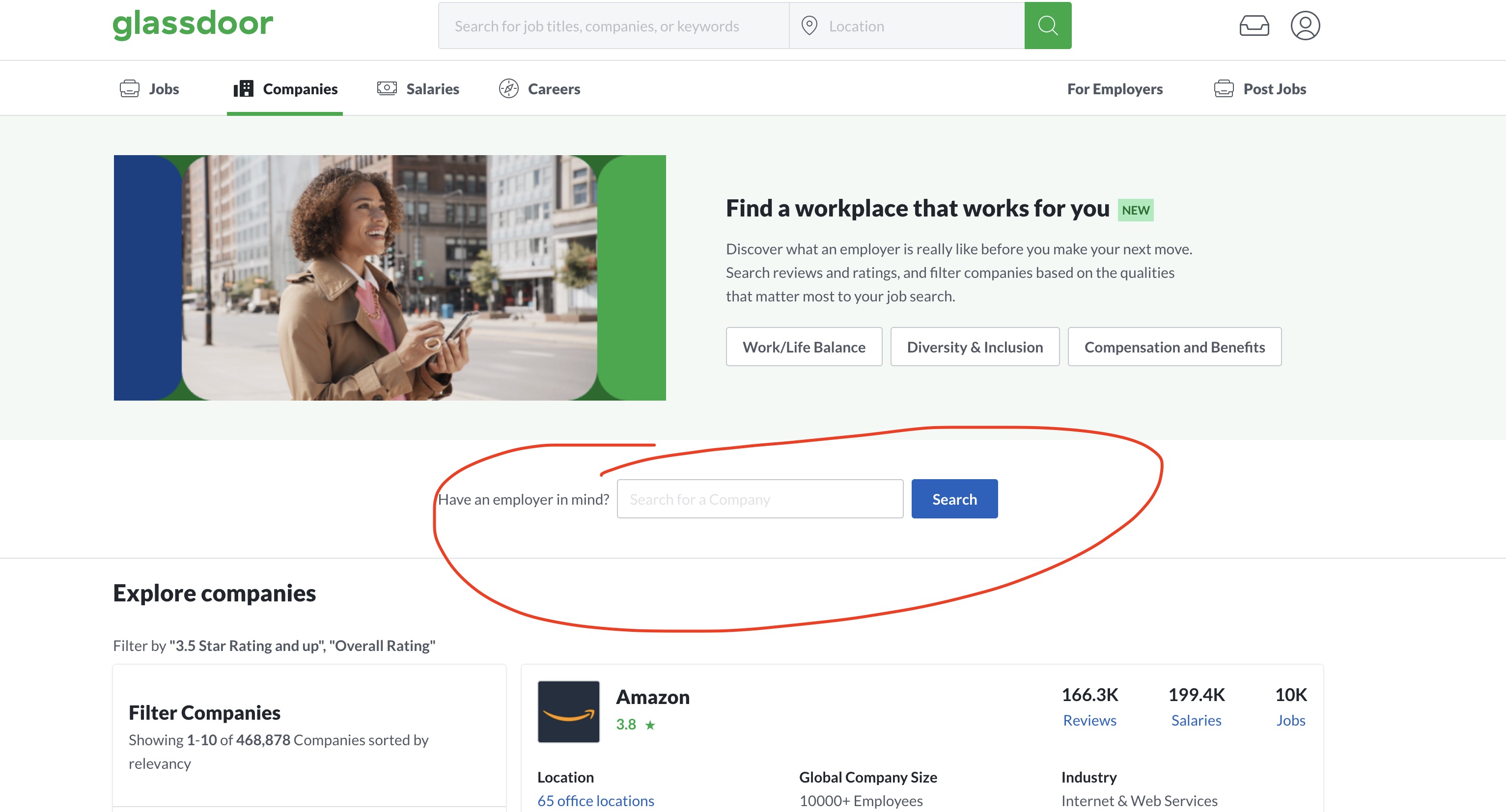
有三个公司名称表，对每一个公司，

在这个网址里<https://www.glassdoor.com/Reviews/index.htm>

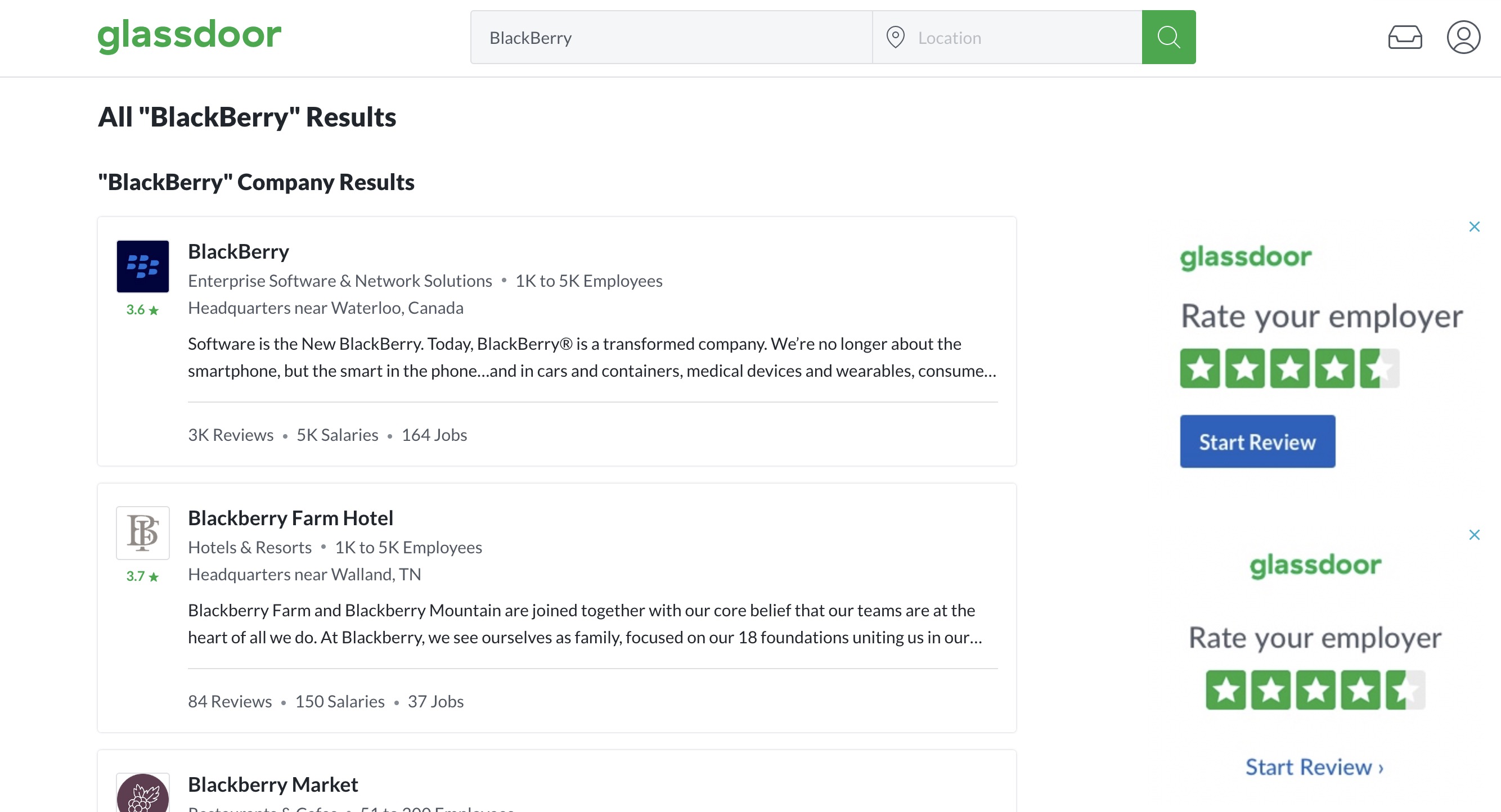
这个搜索框里

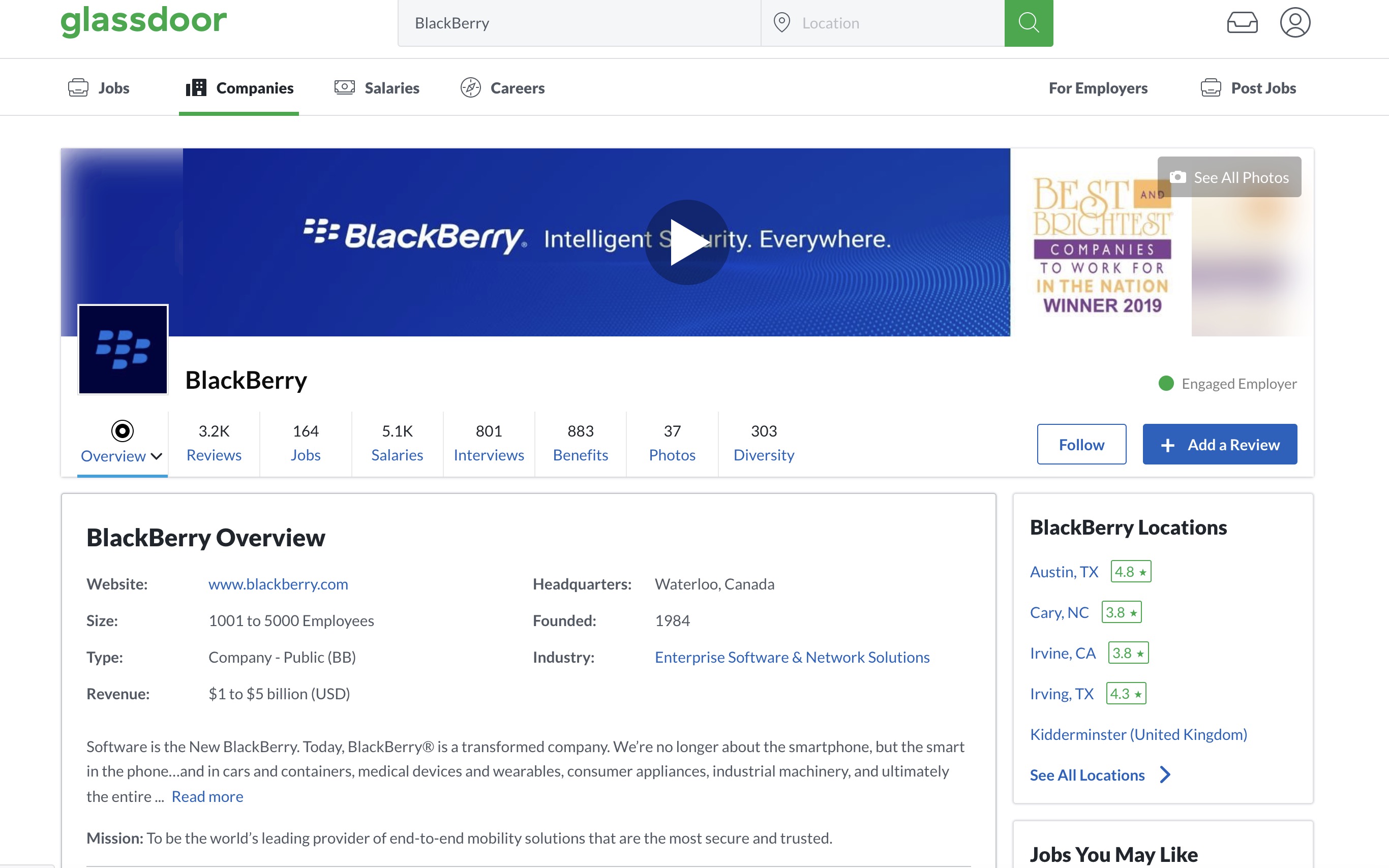


搜索该公司

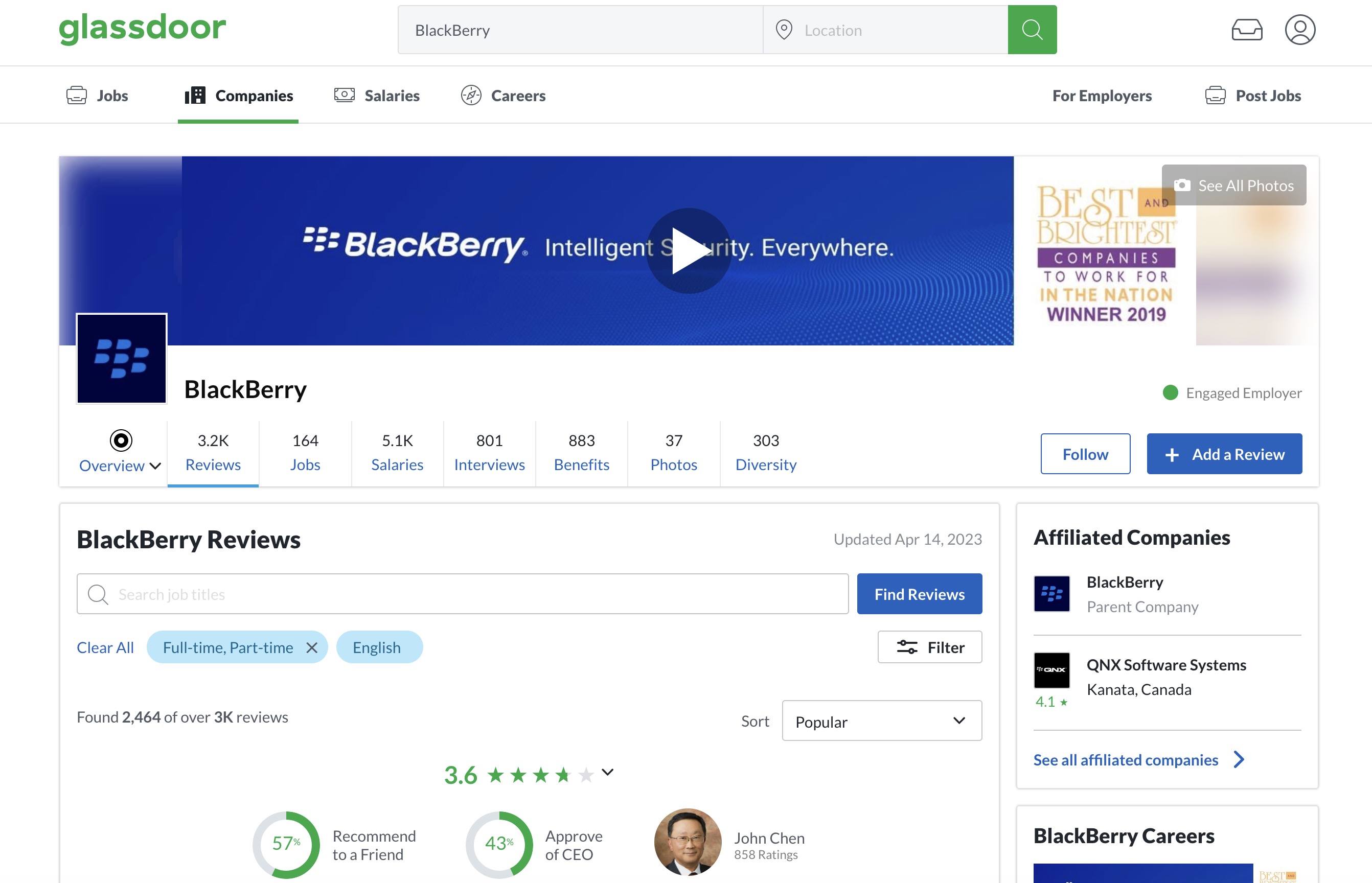
比如，公司叫“BlackBerry”，搜索之后，结果如图：

（此处搜索需要完全一致，比如有一家公司叫TESLA，需要找到这家，而不是Tesla）



点进第一个结果，如图，

点进“reviews”：

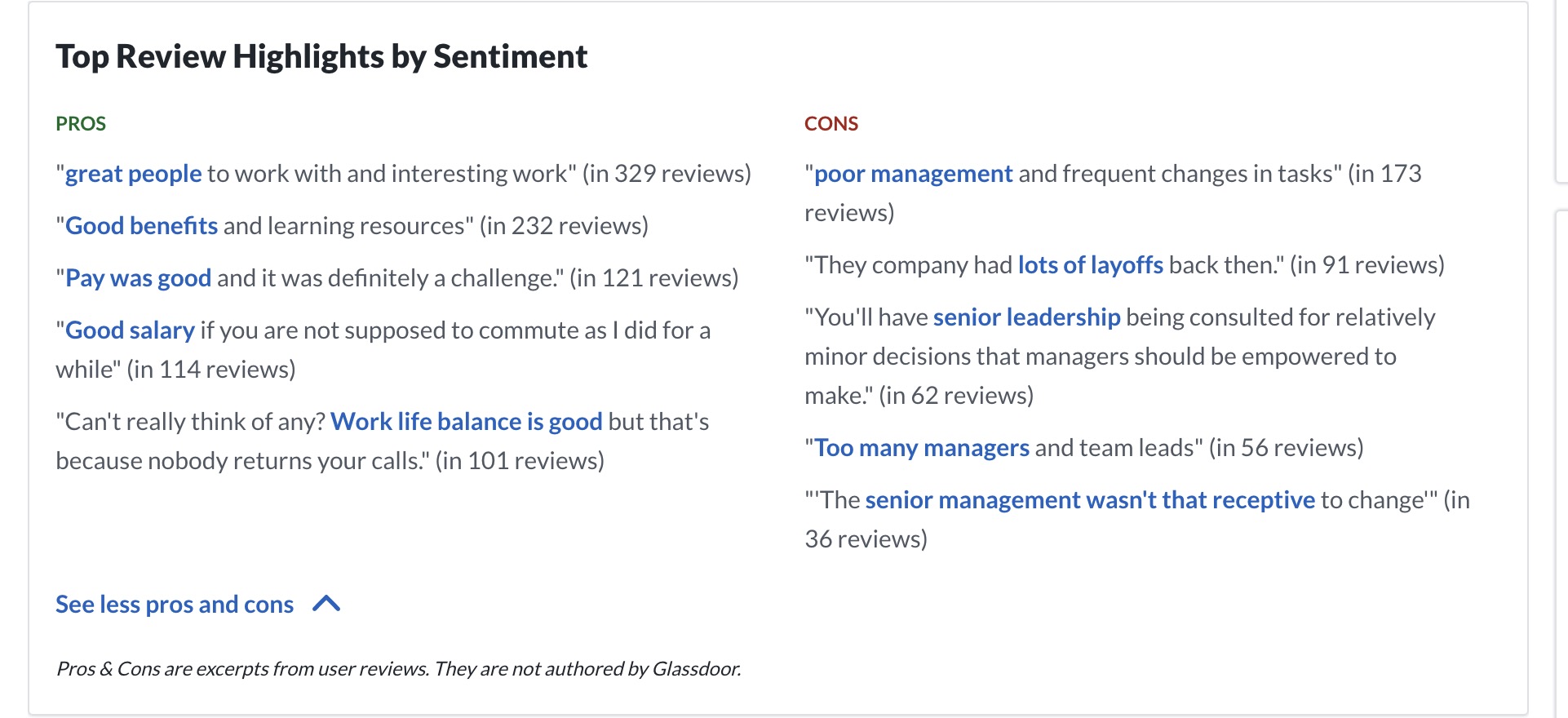


1. 此处爬取 这三个数据： number of reviews(2464), recommend to a friend（57%），approve to CEO (43%）

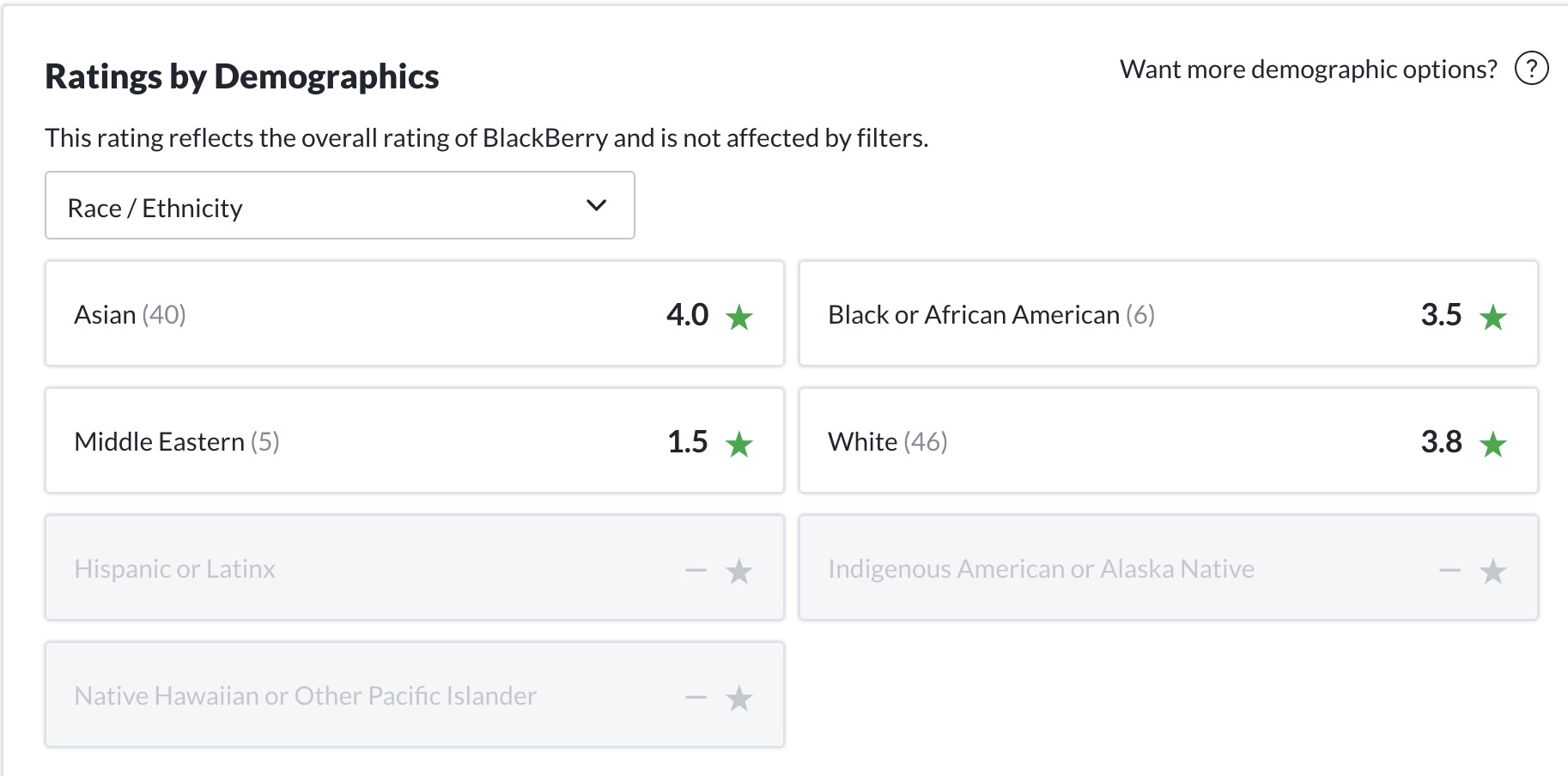
同一页面，向下滑，有这两个部分，

1. 分别爬取圈出来的内容

2.1 pros和cons各五条，所有内容都要爬下来



（这个部分需要手动点击 see more）

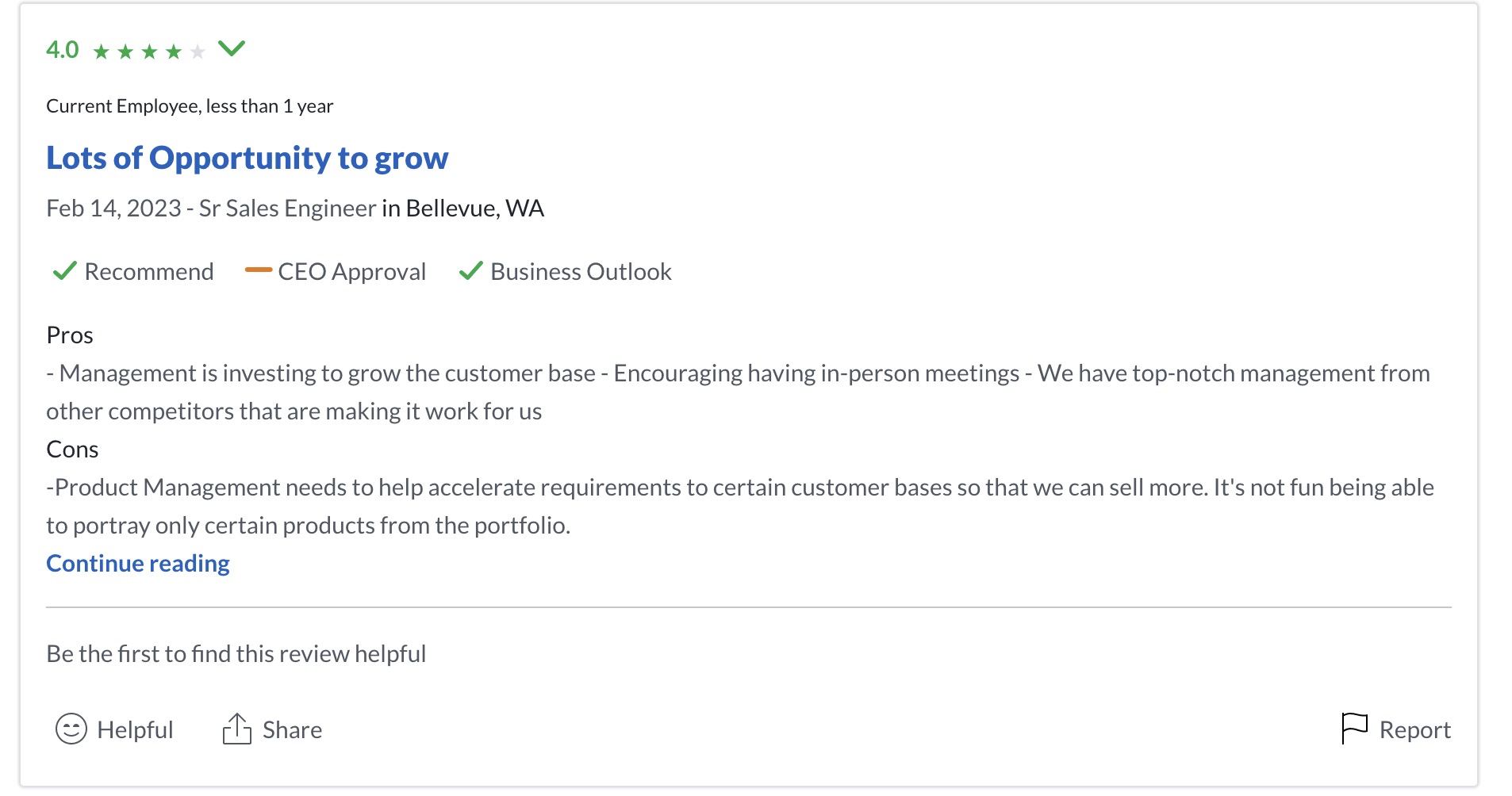
2.2

这个部分要把这些分数都要爬下来,如

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| asin | Black of African American | Middle eastern | white | Hispanic or Latinx | indigenous | Pacific Islander |
| 4.0 | 3.5 | 1.5 | 3.8 |  |  |  |

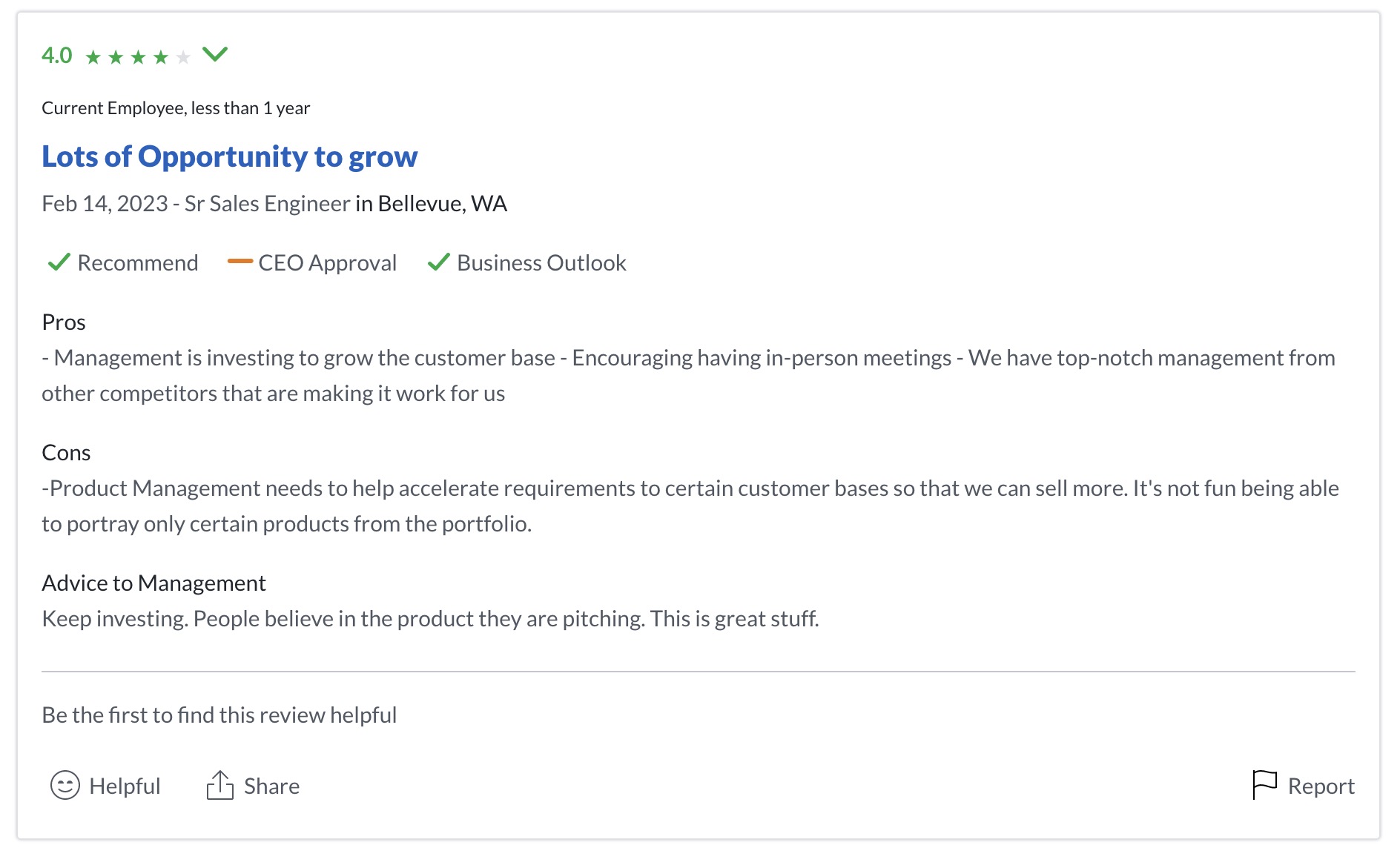
1. 继续向下滑，就是所有的评论，每一个评论如图所示的结构

有的时候会遇到评论显示不全需要点击“continue reading”



点击完如图

Score



helpful

time

position

title

status

对每条评论，爬取红框中的数据

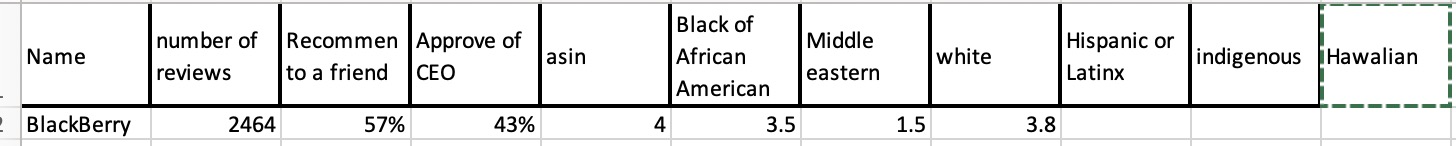
Score, status, title, time, position, recommend, CEO Approval, Business Outlook, pros, cons, advice to management，helpful

（advice to management 这个部分不一定每一条评论都会有，但是点击continue read 后有的话，就要爬）

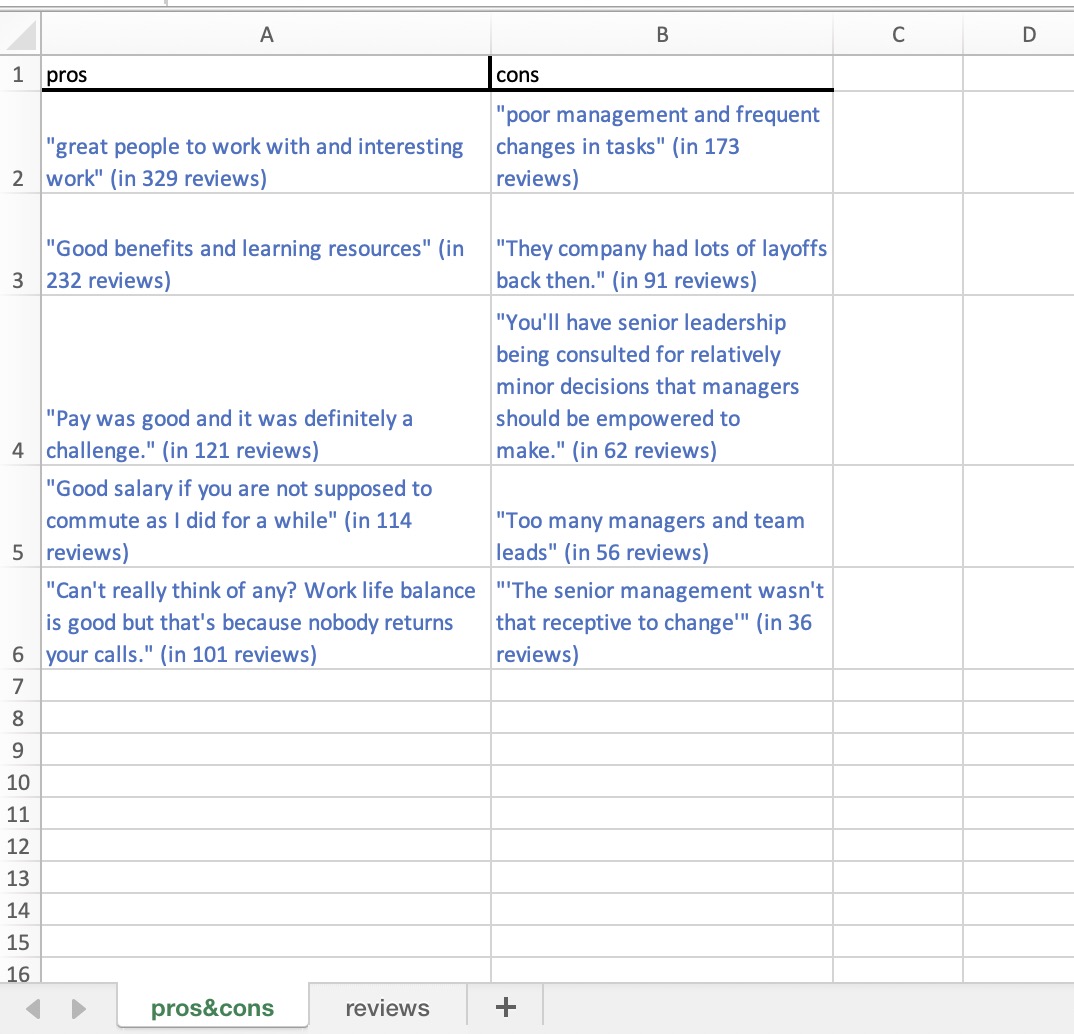
然后爬取所有的评论

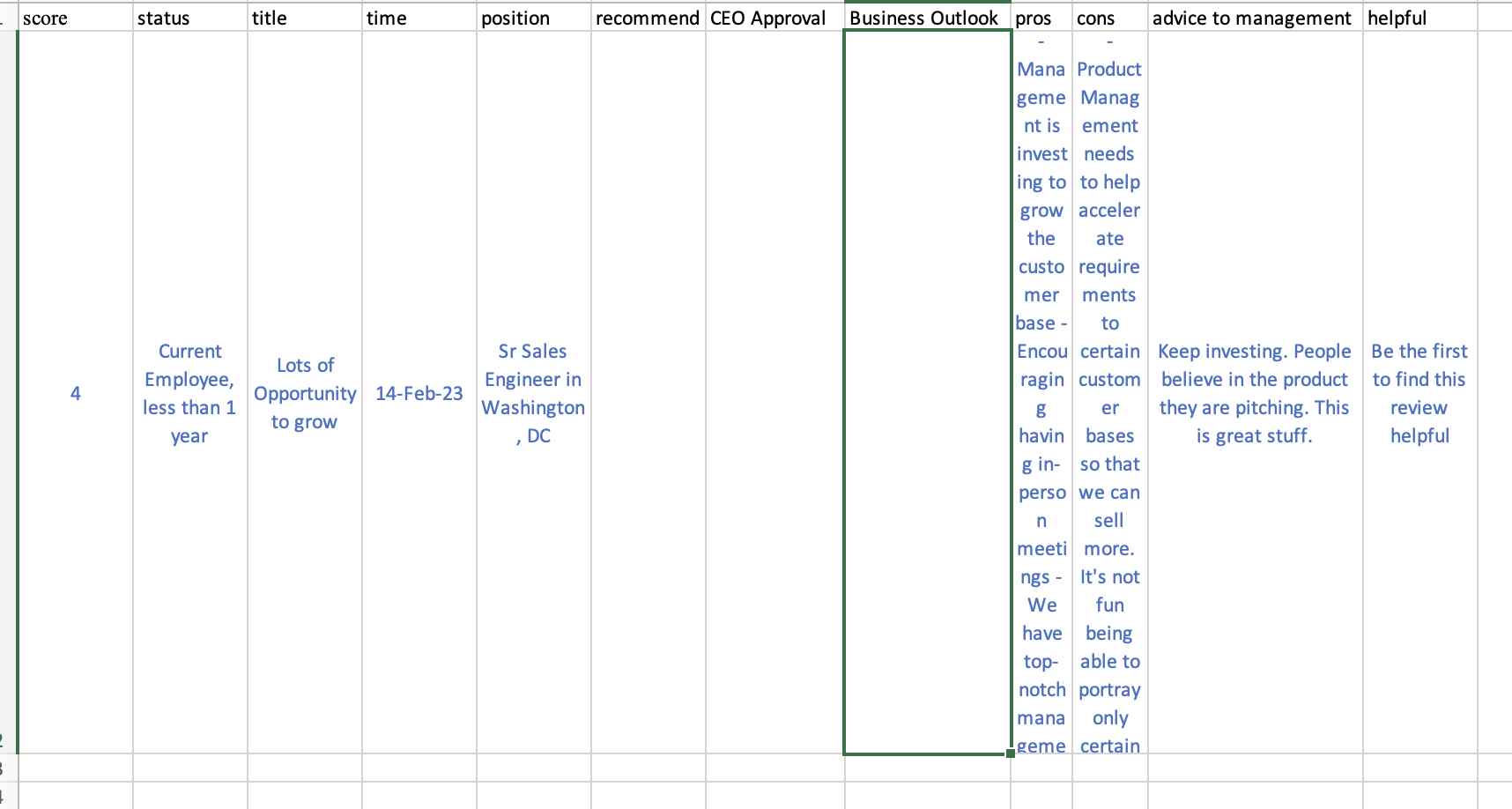
最后的结果如下给出（供参考，可以是不同的表格结构，只要数据都有就行）：

1. 分行业给出各种数值的数据（三个表，公司粒度—>一个公司一行）



1. 每个公司一个excel，sheet1存 pros& cons，sheet2 存所有的reviews





其中 recommend、CEO approval、business Outlook 三列什么表达形式都可以，只要能区分出来就可以