

✓ 收集天猫评论 - 安全座椅，手推车

✓ 爬取网页端评论（评论不全，每个sku约80条）

```

1 # 爬取网页
2 from selenium import webdriver
3 from selenium.webdriver.common.by import By
4 import time
5
6 # 启动浏览器驱动
7 driver = webdriver.Chrome()
8
9 # 打开网页，安全座椅，手推车
10 url = "https://detail.tmall.com/item.htm?_u=t2dmg8j26111&id=698046317552"
11 # url = "https://detail.tmall.com/item.htm?_u=t2dmg8j26111&id=618368232224&spm=a1z0k.1000777.0.0.657837deHpcgv3"
12 driver.get(url)
13
14 time.sleep(30)
15 # 此处先手动登录，后续可用cookie
16 print('login done')
17
18 # 通过class查找元素并点击
19 element = driver.find_element(By.CLASS_NAME, "ShowButton--o4XEG7ih")
20 element.click()
21
22 time.sleep(5)
23 print('click comment button done')
24
25
26
27 # 滚动页面
28 # class="comments--vOMSBfi2 beautify-scroll-bar"
29 div_element = driver.find_element(By.CLASS_NAME, "comments--vOMSBfi2")
30 scroll_script = """
31 var scroll = setInterval(function(){
32     div_element.scrollBy(0,300)
33 }, 3000);
34 """
35 driver.execute_script(scroll_script)
36 # 等待一段时间让页面滚动
37 # 此处先手动滚动
38 time.sleep(60)
39 # 停止滚动
40 driver.execute_script("clearInterval(scroll)")
41 print('scroll to view all comment done')
42
43
44 # 获取页面源代码
45 page_source = driver.page_source
46
47 # 关闭浏览器驱动
48 driver.quit()


1 # 保存网页
2 # with open('./data/safety_seat.html', 'w', encoding='utf-8') as file:
3 #     file.write(page_source)
4 with open('./data/stroller.html', 'w', encoding='utf-8') as file:
5     file.write(page_source)


1 # 解析评论内容
2 from bs4 import BeautifulSoup
3 from pprint import pprint
4 import csv
5
6 # html = 'safety_seat.html'
7 html = 'stroller.html'
8 with open('./data/'+html, 'r', encoding='utf-8') as file:
9     page_source = file.read()
10
11 soup = BeautifulSoup(page_source, 'html.parser')
12 # 找到评论所在的元素
13 comments = soup.find_all(class_='content--FpIOzHeP')
14
15 comments_data = {}
16 contents = []

```

```
17
18 for comment in comments:
19     content = comment.get_text(strip=True)
20     contents.append(content)
21
22 comments_data = {'content':contents}
23 print(len(contents))
24
```

↩ 94

```
1 # 保存评论内容
2 import pandas as pd
3
4 # 保存
5 df = pd.DataFrame(comments_data)
6
7 # file_name = 'comments_safetyseat.csv'
8 file_name = 'comments_stroller.csv'
9 df.to_csv('./data/'+file_name, index=False, encoding='utf-8')
10
11 # 查看
12 df = pd.read_csv('./data/'+file_name, encoding='utf-8')
13 display(df)
```

↩

| | content |
|-----|--|
| 0 | 身边同事们都买的这个，就下单买了，真的很舒适很不错，天天带娃出去溜，娃长大了出去玩也能用，真不错！ |
| 1 | 很好组装，颜值在线，重要的是车身很稳，使用稳定顺滑。宝宝还小还没出门，目前就在家使用，也很愿... |
| 2 | 身边同事们都买的这个，就下单买了，真的很舒适很不错，天天带娃出去溜，娃长大了出去玩也能用，真不错！ |
| 3 | 很好组装，颜值在线，重要的是车身很稳，使用稳定顺滑。宝宝还小还没出门，目前就在家使用，也很愿... |
| 4 | 看了好多测评最终选择了bebebus\n ... |
| ... | ... |
| 89 | 对比了oneya, dearmom后，最终选择的bebebus，颜值高，质量好。今天收到就去遛... |
| 90 | 宝宝刚满月，试了一下，稳定性高，对比了几家还是决定买的这个，一分价钱一分货，推着比较稳👍👍👍👍👍👍 |
| 91 | 非常好用，宝宝坐的很舒服出门必备 |
| 92 | 0天后追评：刚溜一圈回来，宝宝坐着很舒服，一点都不闹 |
| 93 | 颜值高，真是太好看了，操控也很好，喜欢❤️ |

94 rows × 1 columns

✓ 爬取手机端评论（每个sku约几千条）

（后续再爬取）