

```
In [49]: import pandas as pd
import numpy as np
```

```
In [16]: main = pd.read_csv('E:\爬虫\jjwx\作品库所有作品（数据少但是优质）\作品库.csv')
main
```

Out[16]:

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	类型_1	类型_2	类型_3
0	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15400449024	2018-07-24 23:08:19	完结作品	原创	言情	近代现代
1	Zoody	oneauthor.php?authorid=3310540	云边咖啡馆	onebook.php?novelid=5095598	轻松	完结	87661	384190016	2020-08-16 06:46:34	完结半价/包月	原创	言情	近代现代
2	竹已	oneauthor.php?authorid=1938546	难哄	onebook.php?novelid=4001734	轻松	完结	389894	14779252736	2019-01-21 00:29:54	完结作品	原创	言情	近代现代
3	栖见	oneauthor.php?authorid=585107	白日梦我	onebook.php?novelid=3525826	正剧	完结	427269	9028750336	2018-03-16 13:21:32	完结作品	原创	言情	近代现代
4	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15391267840	2018-07-24 23:08:19	VIP作品	原创	言情	近代现代
...	...	...	...	...	...	...	...	...	...	...	...	...	...
61543	千佑 Alleycat	oneauthor.php?authorid=2417786	嫌疑人每天都会把侦探气个半死	onebook.php?novelid=6273634	轻松	连载	61162	8173361	2021-10-18 12:58:04	免费作品	衍生	无CP	近代现代
61544	闲云向晚	oneauthor.php?authorid=2415488	我做妖王迷弟的那些年	onebook.php?novelid=4011416	轻松	连载	163686	23800602	2019-01-27 23:53:35	免费作品	衍生	纯爱	古色古香

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	类型_1	类型_2	类型_3
61545	河鳢的脑花	oneauthor.php?authorid=2213495	【全职/恋与】叶修x白起等你在风里	onebook.php?novelid=3500790	正剧	连载	317075	12327059	2018-02-21 17:10:15	免费作品	衍生	纯爱	近现代
61546	聪明机智菜菜菜	oneauthor.php?authorid=1351301	西游记之九世一生	onebook.php?novelid=2507295	正剧	完结	282775	39535896	2015-07-16 11:56:54	免费作品	衍生	纯爱	古色古香
61547	残阳落暖	oneauthor.php?authorid=642847	我有特殊的侧写技巧	onebook.php?novelid=3554034	正剧	暂停	119208	19726530	2018-04-10 22:56:55	免费作品	衍生	纯爱	近现代

61548 rows × 14 columns

In [17]:

author = pd.read\_csv('E:\爬虫\jjwx\作品库所有作品（数据少但是优质）\作者信息.csv')  
author

	a_作者	b_被收藏数	c_最近更新作品	d_作品状态	e_作品字数	f_最后更新时间	g_作者所发送红包数	h_小说完本数	i_小说连载数	j_小说暂停数
0	Zoody	15070	《我的神他心比石硬》	完结	143517	2022-02-15 17:34:52	27	7	10	0
1	栖见	253543	《桃枝气泡》	完结	394389	2021-04-12 05:01:00	23305	29	4	0
2	木苏里	721056	《不见上仙三百年》	连载	309531	2022-03-12 00:55:29	193238	13	6	1
3	唧唧的猫	140448	《是心跳说谎》	完结	330722	2021-08-10 20:03:51	11880	6	2	0
4	竹已	355330	《折月亮》	连载	318390	2022-03-12 21:59:59	23944	8	1	0
...	...	...	...	...	...	...	...	...	...	...
19239	冬蝉鸣夏	192	《cos旧剑后脚踩咒术界高层》	连载	143640	2022-02-16 03:21:01	4	0	7	0
19240	上秋未晴	313	《[综]诱捕三日月》	连载	203646	2021-08-06 20:20:20	47	1	7	0
19241	逆风or随落	44	《斗罗之封宇终始》	完结	143531	2017-03-26 10:00:52	0	3	0	1
19242	番茄通心粉	113	《npc在恋爱游戏花式be后》	连载	498982	2022-03-13 18:00:00	5	3	3	0

	a_作者	b_被收藏数	c_最近更新作品	d_作品状态	e_作品字数	f_最后更新时间	g_作者所发送红包数	h_小说完本数	i_小说连载数	j_小说暂停数
19243	摩卡滋味	3672	《直播异世界种田的日子》	连载	329540	2021-12-26 23:40:00	29	10	5	1

19244 rows × 10 columns

In [18]:

```
main_author = pd.merge(main, author, on='a_作者', how='left')
main_author
```

Out[18]:

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	...	类型_4	b_被浏览
0	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15400449024	2018-07-24 23:08:19	完结作品	...	爱情	35533
1	Zoody	oneauthor.php?authorid=3310540	云边咖啡馆	onebook.php?novelid=5095598	轻松	完结	87661	384190016	2020-08-16 06:46:34	完结半价/包月	...	爱情	1507
2	竹已	oneauthor.php?authorid=1938546	难哄	onebook.php?novelid=4001734	轻松	完结	389894	14779252736	2019-01-21 00:29:54	完结作品	...	爱情	35533
3	栖见	oneauthor.php?authorid=585107	白日梦我	onebook.php?novelid=3525826	正剧	完结	427269	9028750336	2018-03-16 13:21:32	完结作品	...	爱情	25354
4	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15391267840	2018-07-24 23:08:19	VIP作品	...	爱情	35533
...	...	...	...	...	...	...	...	...	...	...	...	...	...
61543	千佑Alleycat	oneauthor.php?authorid=2417786	嫌疑人每天都会把侦探气个半死	onebook.php?novelid=6273634	轻松	连载	61162	8173361	2021-10-18 12:58:04	免费作品	...	轻小说	124
61544	闲云向晚	oneauthor.php?authorid=2415488	我做妖王迷弟的那些年	onebook.php?novelid=4011416	轻松	连载	163686	23800602	2019-01-27 23:53:35	免费作品	...	东方衍生	3

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	...	类型_4	b_初藏
61545	河鳢的脑花	oneauthor.php?authorid=2213495	【全职/恋与】叶修x白起等你在风里	onebook.php?novelid=3500790	正剧	连载	317075	12327059	2018-02-21 17:10:15	免费作品	...	东方衍生	1
61546	聪明机智菜菜菜	oneauthor.php?authorid=1351301	西游记之九世一生	onebook.php?novelid=2507295	正剧	完结	282775	39535896	2015-07-16 11:56:54	免费作品	...	古典衍生	164
61547	残阳落暖	oneauthor.php?authorid=642847	我有特殊的侧写技巧	onebook.php?novelid=3554034	正剧	暂停	119208	19726530	2018-04-10 22:56:55	免费作品	...	西方衍生	149

61548 rows × 23 columns

In [19]:

```
piece = pd.read_csv('E:\爬虫\jjwx\作品库所有作品（数据少但是优质）\文章信息.csv')
piece['piece_url'] = piece['piece_url'].apply(lambda x: x[22:].strip())
piece = piece.rename(columns={'piece_url': 'b_作品_url'})
piece
```

Out[19]:

	b_作品_url	作品视角	作品风格	全文字数	前进一名所需地雷数	总书评数	总共地雷数量	总点击量	文章名称	文章当前被收藏数	文章积分	文章类型	文章进度	签约状态
0	onebook.php?novelid=2595385	女主	正剧	698356	89	88954	48892	14047236	有匪	409775	6455118848	原创-言情-架空历史-武侠	完结	已签约

[illegible]

	b_作品_url	作品 视角	作品 风格	全文 字数	前进 一名 所需 地雷 数	总书 评数	总 共 地雷 数量	总 点 击 量	文章 名称	文章 当 前 被 收 藏 数	文章 积 分	文章 类 型	文章 进 度	总 状 态
41297	onebook.php?novelid=237304	NaN	正 剧	1297466	1	9674	5	3445218	风 *****)	2340	159107120	衍生-未知-古色古香-东方衍生	完结	已完结
41298	onebook.php?novelid=5619492	主攻	正 剧	58139	1	513	0	382106	当一梦江湖的奶爸来到山河令	1950	6340120	衍生-纯爱-古色古香-其他衍生	连载	未完结
41299	onebook.php?novelid=2966014	不明	正 剧	542964	1	1707	73	3725932	[综]红楼梦死亡不完全手册	1950	39957068	衍生-无CP-幻想未来-古典衍生	完结	已完结
41300	onebook.php?novelid=2520185	主受	正 剧	143531	1	222	4	2057553	斗罗之封宇终始	1934	7821952	衍生-纯爱-架空历史-东方衍生	完结	未完结

	b_作品_url	作品 视角	作品 风格	全文 字数	前进 一名 所需 地雷 数	总书 评数	总 共 地 雷 数	总 点 击 量	文章 名 称	文章 当 前 被 收 藏 数	文章 积 分	文章 类 型	文章 进 度	总 状 态
41301	onebook.php?novelid=491678	主受	轻松	46363	1	1572	5	640879	[HP] 懒散度日	1949	8003392	衍生-纯爱-近代-现代-西方衍生	连载	未完结

41302 rows × 16 columns

In [37]:

```
data = pd.merge(main_author, piece, on='b_作品_url', how='left')
data
```

Out[37]:

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品 库 类 型	...	总 共 地 雷 数 量	
0	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15400449024	2018-07-24 23:08:19	完结作品	...	16878.0	2
1	Zoody	oneauthor.php?authorid=3310540	云边咖啡馆	onebook.php?novelid=5095598	轻松	完结	87661	384190016	2020-08-16 06:46:34	完结半价/包月	...	234.0	

	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	...	总共地雷数量
2	竹已	oneauthor.php?authorid=1938546	难哄	onebook.php?novelid=4001734	轻松	完结	389894	14779252736	2019-01-21 00:29:54	完结作品	...	33901.0
3	栖见	oneauthor.php?authorid=585107	白日梦我	onebook.php?novelid=3525826	正剧	完结	427269	9028750336	2018-03-16 13:21:32	完结作品	...	17660.0
4	竹已	oneauthor.php?authorid=1938546	偷偷藏不住	onebook.php?novelid=3676002	轻松	完结	372127	15391267840	2018-07-24 23:08:19	VIP作品	...	16878.0
...	...	...	...	...	...	...	...	...	...	...	...	...
61543	千佑 Alleycat	oneauthor.php?authorid=2417786	嫌疑人每天都会把侦探气个半死	onebook.php?novelid=6273634	轻松	连载	61162	8173361	2021-10-18 12:58:04	免费作品	...	0.0
61544	闲云向晚	oneauthor.php?authorid=2415488	我做妖王迷弟的那些年	onebook.php?novelid=4011416	轻松	连载	163686	23800602	2019-01-27 23:53:35	免费作品	...	17.0



	a_作者	a_作者_url	b_作品	b_作品_url	d_风格	e_进度	f_字数	g_作品积分	h_发表时间	作品库类型	...	总共地雷数量
61545	河鱔的脑花	oneauthor.php?authorid=2213495	【全职/恋与】叶修x白起等你在风里	onebook.php?novelid=3500790	正剧	连载	317075	12327059	2018-02-21 17:10:15	免费作品	...	13.0
61546	聪明机智菜菜菜	oneauthor.php?authorid=1351301	西游记之九世一生	onebook.php?novelid=2507295	正剧	完结	282775	39535896	2015-07-16 11:56:54	免费作品	...	205.0
61547	残阳落暖	oneauthor.php?authorid=642847	我有特殊的侧写技巧	onebook.php?novelid=3554034	正剧	暂停	119208	19726530	2018-04-10 22:56:55	免费作品	...	20.0

In [38]:

Out[38]:

[illegible]

	a_作者	a_作者_url	b_作品_url	h_发表时间	作品库类型	b_被收藏数	c_最近更新作品	d_作品状态	e_作品字数	f_最后更新时间	...	总霍
61543	千佑 Alleycat	oneauthor.php?authorid=2417786	onebook.php?novelid=6273634	2021-10-18 12:58:04	免费作品	1242.0	《嫌疑人每天都会把侦探气个半死》	连载	61162.0	2022-03-04 01:10:18	...	
61544	闲云向晚	oneauthor.php?authorid=2415488	onebook.php?novelid=4011416	2019-01-27 23:53:35	免费作品	36.0	《我做妖王迷弟的那些年》	连载	163686.0	2019-03-26 23:01:01	...	
61545	河鳢的 脑花	oneauthor.php?authorid=2213495	onebook.php?novelid=3500790	2018-02-21 17:10:15	免费作品	18.0	《【全职/恋与】叶修x白起等你在风里》	连载	317075.0	2021-06-01 15:51:07	...	
61546	聪明机智菜菜菜	oneauthor.php?authorid=1351301	onebook.php?novelid=2507295	2015-07-16 11:56:54	免费作品	1644.0	《久*****】》	连载	156901.0	2021-01-02 00:53:53	...	

	a_作者	a_作者_url	b_作品_url	h_发表时间	作品库类型	b_被收藏数	c_最近更新作品	d_作品状态	e_作品字数	f_最后更新时间	...	总篇数
61547	残阳落暖	oneauthor.php?authorid=642847	onebook.php?novelid=3554034	2018-04-10 22:56:55	免费作品	1495.0	《[足球]冠军之巅》	连载	587347.0	2022-01-04 23:22:55	...	

61548 rows × 29 columns

```
In [44]: data.columns = ['a_作者', 'author_url','piece_url', 'b_发表时间', 'c_作品库类型', 'a_被收藏数',
                        'a_最后更新时间', 'a_作者所发送红包数', 'a_小说完本数', 'a_小说连载数', 'a_小说暂停',
                        'b_前进一名所需地雷数', 'b_总书评数', 'b_总共地雷数量', 'b_总点击量', 'b_文章名称',
                        'b_文章进度', 'b_签约状态', 'b_营养液数', 'b_霸王票全站排行']
```

```
In [45]: data
```

Out[45]:

	a_作者	author_url	piece_url	b_发表时间	c_作品库类型	a_被收藏数	a_最近更新作品	a_作品状态	a_作品字数	a_最后更新时间	...	b_总篇数
0	竹已	oneauthor.php?authorid=1938546	onebook.php?novelid=3676002	2018-07-24 23:08:19	完结作品	355330.0	《折月亮》	连载	318390.0	2022-03-12 21:59:59	...	16
1	Zoody	oneauthor.php?authorid=3310540	onebook.php?novelid=5095598	2020-08-16 06:46:34	完结半价/包月	15070.0	《我的神他心比石硬》	完结	143517.0	2022-02-15 17:34:52	...	

	a_作者	author_url	piece_url	b_发表时间	c_作品库类型	a_被收藏数	a_最近更新作品	a_作品状态	a_作品字数	a_最后更新时间	...	b_地址
2	竹已	oneauthor.php?authorid=1938546	onebook.php?novelid=4001734	2019-01-21 00:29:54	完结作品	355330.0	《折月亮》	连载	318390.0	2022-03-12 21:59:59	...	33
3	栖见	oneauthor.php?authorid=585107	onebook.php?novelid=3525826	2018-03-16 13:21:32	完结作品	253543.0	《桃枝气泡》	完结	394389.0	2021-04-12 05:01:00	...	17
4	竹已	oneauthor.php?authorid=1938546	onebook.php?novelid=3676002	2018-07-24 23:08:19	VIP作品	355330.0	《折月亮》	连载	318390.0	2022-03-12 21:59:59	...	16
...	...	...	...	...	...	...	...	...	...	...	...	...
61543	千佑 Alleycat	oneauthor.php?authorid=2417786	onebook.php?novelid=6273634	2021-10-18 12:58:04	免费作品	1242.0	《嫌疑人每天都会把侦探气个半死》	连载	61162.0	2022-03-04 01:10:18	...	
61544	闲云向晚	oneauthor.php?authorid=2415488	onebook.php?novelid=4011416	2019-01-27 23:53:35	免费作品	36.0	《我做妖王迷弟的那些年》	连载	163686.0	2019-03-26 23:01:01	...	

	a_作者	author_url	piece_url	b_发表时间	c_作品库类型	a_被收藏数	a_最近更新作品	a_作品状态	a_作品字数	a_最后更新时间	...	b_地址
61545	河鳢的脑花	oneauthor.php?authorid=2213495	onebook.php?novelid=3500790	2018-02-21 17:10:15	免费作品	18.0	《【全职/恋与】叶修x白起 等你在风里》	连载	317075.0	2021-06-01 15:51:07	...	
61546	聪明机智菜菜菜	oneauthor.php?authorid=1351301	onebook.php?novelid=2507295	2015-07-16 11:56:54	免费作品	1644.0	《久*****】》	连载	156901.0	2021-01-02 00:53:53	...	
61547	残阳落暖	oneauthor.php?authorid=642847	onebook.php?novelid=3554034	2018-04-10 22:56:55	免费作品	1495.0	《[足球]冠军之巅》	连载	587347.0	2022-01-04 23:22:55	...	

61548 rows × 29 columns

```
In [46]: # 保存至文件
data.to_csv('E:\爬虫\jjwx\作品库所有作品（数据少但是优质）\作品库爬取.csv', index=False)

In [ ]:

In [ ]:
```